



HAL
open science

Extreme gradient boosting trees with efficient Bayesian optimization for profit-driven customer churn prediction

Zhenkun Liu, Ping Jiang, Koen W. de Bock, Jianzhou Wang, Lifang Zhang,
Xinsong Niu

► To cite this version:

Zhenkun Liu, Ping Jiang, Koen W. de Bock, Jianzhou Wang, Lifang Zhang, et al.. Extreme gradient boosting trees with efficient Bayesian optimization for profit-driven customer churn prediction. *Technological Forecasting and Social Change*, 2024, 198 (January 2024), pp.122945. 10.1016/j.techfore.2023.122945 . hal-04273578

HAL Id: hal-04273578

<https://hal.science/hal-04273578v1>

Submitted on 7 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0
International License

Extreme Gradient Boosting Trees with Efficient Bayesian Optimization for Profit-Driven Customer Churn Prediction

Zhenkun Liu¹, Ping Jiang^{2*}, Koen W. De Bock³, Jianzhou Wang⁴, Lifang Zhang⁵ and Xinsong Niu⁶

¹School of Management, Nanjing University of Posts and Telecommunications, No 66 Xinmofan Road, Gulou District, Nanjing, Jiangsu Province 210023, China

²School of Statistics, Dongbei University of Finance and Economics, No. 217, Jianshan Road, Shahekou District, Dalian, Liaoning Province 116025, China

³Audencia Business School, 8 Route de la Jonelière, F-44312 Nantes, France

⁴Institute of Systems Engineering, Macau University of Science and Technology, Taipa Street, Macao 999078, China

⁵School of Finance, Nanjing University of Finance and Economics, No.3 Wenyuan Road, Qixia District, Nanjing, Jiangsu Province 210023, China

⁶Academy of Mathematics and Systems Science, Chinese Academy of Sciences, No.55 Zhongguancun East Road, Haidian District, Beijing 100190, China

Preprint version

Cite as:

Liu, Z., Jiang, P., Bock, K. W. De, Wang, J., Zhang, L., & Niu, X. (2024). Technological Forecasting & Social Change Extreme gradient boosting trees with efficient Bayesian optimization for profit-driven customer churn prediction. *Technological Forecasting & Social Change*, 198(October 2023), 122945.

<https://doi.org/10.1016/j.techfore.2023.122945>

<https://www.sciencedirect.com/science/article/pii/S0040162523006303>

* Corresponding author: Ping Jiang, Tel.: +86 17866540741

E-mail addresses: Zhenkun Liu (zhenkunliudufe@163.com);

Ping Jiang (pjiang@dufe.edu.cn);

Koen W. De Bock (kdebock@audencia.com);

Jianzhou Wang (jianzhouwudufe@163.com);

Lifang Zhang (lifangzhang1106@126.com);

Xinsong Niu (xinsongniu@126.com)

Acknowledgement

This research was supported by the Major Program of National Fund of Philosophy and Social Science of China (Grant No. 19ZDA120).

Extreme Gradient Boosting Trees with Efficient Bayesian Optimization for Profit-Driven Customer Churn Prediction

Abstract

Customer retention campaigns increasingly rely on predictive analytics to identify potential churners within the customer base. Traditionally, customer churn prediction has depended on binary classifiers, often optimized for accuracy-based performance measures. However, there is a growing consensus that this approach may not always fulfill the critical business objective of profit maximization, as it overlooks the costs of misclassification and the benefits of accurate classification. To conduct profit-driven customer churn prediction, this study adopts extreme gradient boosting trees. The class weights and other hyper-parameters of these trees are optimized using Bayesian methods based on the maximum profit criterion. Empirical analyses are carried out using real datasets obtained from service providers in multiple markets. The empirical results demonstrate that the proposed model yields significantly higher profits compared to benchmark models. Both Bayesian optimization and the adjustment of class weights contribute to an enhanced model profitability. Furthermore, when optimizing multiple hyperparameters, the computational cost of model optimization is significantly reduced compared to exhaustive grid search. Additionally, we showcase the robustness of the proposed model through a sensitivity analysis, employing Bayesian optimization. With the proposed model, marketing managers can design targeted marketing plans to retain customer groups with a higher likelihood of churning.

Keywords: Bayesian optimization; Extreme gradient boosting tree; Profit-driven customer churn prediction; Sensitivity analysis

1 Introduction

Establishing a robust relationship with customers is imperative for any business functioning within a dynamic and competitive market and macro-environment (Coussement et al., 2017). Given the fungibility and limited differentiation among service providers in competitive markets, customers tend to readily compare and switch between service providers or services. Customers who choose to leave

are commonly referred to as churned customers (Óskarsdóttir et al., 2017). As a vital component of *customer relationship management (CRM)*¹, proactive customer retention management concentrates on preserving existing customers, who exhibit lower churn rates (Dawes Farquhar, 2004; Reinartz and Kumar, 2002), are less susceptible to competitive marketing efforts (Colgate et al., 1996), generate more sales, require fewer resources (Ganesh et al., 2000), and can attract additional customers through positive word of mouth (Ganesh et al., 2000). Conversely, customers who churn are likely to influence the choices and opinions of customers within their social network, in line with social ties and homophily theories, potentially leading to further instances of churn (Nitzan and Libai, 2011). Furthermore, churn necessitates that companies invest in acquiring new customers, a cost that can be up to five to six times higher than retaining an existing client (Gandomi and Zolfaghari, 2013; Ganesh et al., 2000; Van den Poel and Larivière, 2004).

Customer retention can greatly benefit from *Customer Churn Prediction (CCP)*, which enables companies to identify customers with a high inclination to switch to competitors (Ganesh et al., 2000). When deploying churn prediction to support customer retention, the effectiveness of campaigns, and thus their profitability, primarily depends on the predictive accuracy of the underlying models. The CCP literature is primarily dedicated to exploring alternative strategies for achieving higher levels of model accuracy through various means. The first stream of literature investigates the added value of new features or novel data sources (e.g., Óskarsdóttir et al., 2017; De Caigny, Coussement, De Bock, & Lessmann, 2020). A second common strategy involves variations and improvements in the classifiers used to construct CCP models (e.g., Pekel Ozmen & Ozcan, 2022). Nowadays, there is a consensus that model hyperparameters can significantly impact model performance and thus should be optimized. This optimization can be achieved through exhaustive methods like grid search or more intelligent approaches like genetic algorithms. Finally, there exists a strong tendency in CCP literature to employ alternative, often profit-based evaluation metrics for selecting, optimizing, and evaluating models. In this study, we present a combined strategy that focuses on the latter three dimensions (i.e., classifier selection, model optimization, and adoption of evaluation metrics) in pursuit of improved predictive accuracy and computational efficiency (e.g., Stripling et al. 2018).

¹ List of abbreviations is shown in Table A1 in Appendix.

A noteworthy category of classifiers with a convincing track record in business analysis are *ensemble learners*, models that aggregate predictions of multiple constituent models (Jabeur et al., 2021). In recent years, *extreme gradient boosting tree* (XGBT) (Chen and Guestrin, 2016), is a particularly noteworthy ensemble learner that enjoyed increasing attention in various domains (Al-Shatnwai and Faris, 2020; Hajek et al., 2022; Liu et al., 2023; Senthana et al., 2021). XGBT conducts a second-order Taylor expansion on the loss function using the gradient boosting decision tree method (Chehreh Chelgani et al., 2021). Relative to other gradient boosting techniques, XGBT could select a strong classification model among a series of poor classification models. Noteworthy advantages of XGBT are the following (Luo et al., 2021): (1) they validly deal with missing values; (2) they reduce overfitting; and (3) parallel and distributed calculation enhance computational efficiency. Moreover, its objective is to use a gradient descent optimization algorithm and arbitrary differentiable loss functions to achieve the minimum of loss function via involving poor learners.

Confirming findings in other domains, experiments in multiple CCP studies have demonstrated the competitive nature of XGBT in CCP models. However, while it is nowadays generally recognized that model hyperparameters can have an impact on the predictive accuracy of a CCP classifier and thus should be optimized, we observe that a majority of XGBT applications in CCP forego on parameter optimization. Moreover, in general this optimization involves an exhaustive grid search despite more intelligent and computationally efficient approaches. Furthermore, for the remaining approaches that do implement optimization, in all cases this involved optimizing profit-agnostic performance measure such as accuracy, F1-measure or the *area under the receiver operating characteristics curve* (AUC). In our study, we aim to address these limitations and propose a profit-driven CCP model (BO-XGBT) based on XGBT, and *Bayesian optimization* (BO). To conduct profit-driven CCP, XGBT models are optimized by BO based on the *expected maximum profit criterion* (EMPC). Moreover, besides a set of conventional hyperparameter candidates found in other studies we deliberately opt to optimize the *class weight* parameter of XGBT, which manipulates the relative importance given to minority class observations. This parameter essentially allows changing the loss function minimized during the XGBT training process and is thus especially suitable for optimizing models in terms of profit. To validate the proposed BO-XGBT, experiments on eight datasets originating from customer services companies (in multiple industries) are used to validate BO-XGBT and demonstrate how the particular combination of Bayesian hyperparameter optimization and XGBT is superior to XGBT with hyperparameter optimization based

on grid search on the one hand, and well-established benchmark classifiers on the other, both in terms of predictive performance as computational efficiency.

In customer churn prediction, the proposed BO-XGBT can bring higher profits for enterprises to predict churners. The manager responsible for CRM should pay more attention to potential churners. Once the proposed BO-XGBT identifies customers as potential churners, the marketing manager should immediately contact these customers and provide them with attractive marketing concessions, such as new products or more favorable prices, to prevent churners (Lee et al., 2011). Therefore, based on the proposed BO-XGBT, marketing managers can design targeted marketing plans at the macro level to retain customer groups with churn probability. From a micro perspective, the proposed BO-XGBT helps to identify potential churners and take preventive measures to avoid churners (De Caigny et al., 2020; Lu et al., 2014).

Against the backdrop of technological forecasting and social changes, we recognize the potential for applying specific technologies in the field of customer churn prediction to gain deeper insights and accuracy. This paper aims to explore the utilization of specific technological tools for predicting and addressing customer churn issues, providing novel insights and methodologies for the audience interested in technological forecasting and social changes. Our research provides explicit technology focus in the following areas: 1) Enhancing the ensemble model XGBT for customer churn prediction using specific hyperparameter optimization techniques (Bayesian optimization). 2) Highlighting the application potential of the profit-driven BO-XGBT framework in practical business decision-making, contributing to long-term enterprise development and technological transformation. 3) Providing practitioners with targeted customer retention strategies to improve retention rates and enhance the competitive advantage of businesses.

The remainder of our study is organized below: Section 2 introduces related works of CCP. The methods involved in the proposed BO-XGBT are listed in Section 3. Section 4 establishes the details of experimental validation. Experimental results are discussed in Section 5, and Section 6 is the conclusion.

2 Related Work

This section discusses two specific sub-streams of customer churn prediction literature upon which this study narrowly relies: (i) profit-driven customer churn prediction (Section 2.1), and (ii) churn prediction using extreme gradient boosting trees (Section 2.2).

2.1 Profit-Driven Customer Churn Prediction

To better evaluate the potential profitability of a classification model deployed for customer churn prediction, some scholars proposed adopting profit-based performance indicators criteria (Verbeke et al., 2012; Verbraken et al., 2014, 2013). For example, Verbeke et al. (2012) proposed the *maximum profit criterion* that obtains the maximum value of average classification profit by optimizing classification threshold. Then Verbraken et al. (2013) developed a probabilistic variant, i.e. the EMPC from maximum profit criterion. EMPC accommodates uncertainty with respect to the probability that a would-be churner could be effectively.

Initially, the goal of the profit maximization criterion and EMPC was to select the optimal CCP model by evaluating the profitability of classifiers in a post-hoc fashion. Subsequently, researchers started using the profit metrics as guiding target in the classifier training process to build profit-driven CCP models. For example, Maldonado et al. (2015) developed a profit-driven feature selection method based on support vector machines, which maximizes the CCP profit by removing unrelated features. Besides, Stripling et al. proposed a profit-driven logistic regression model based on genetic algorithms that optimize its coefficients for the goal of profit maximization (Stripling et al., 2018). Along the same line, Höppner et al. (2020) proposed profit-driven decision tree based on evolutionary algorithm, in which the training process is again driven by profit indicators.

Recently, logistic regression (Stripling et al., 2018), least absolute shrinkage and selection operator (Maldonado et al., 2020) or decision tree (Höppner et al., 2020) have been developed as profit-driven CCP models. However, only single models are developed as profit-driven models, and profit-driven ensemble models have not been developed. Inspired by Stripling et al. (2018), Maldonado et al. (2020), and Höppner et al. (2020), we develop a profit-driven ensemble model to maximize EMPC value in its training process.

2.2 Customer Churn Prediction with Extreme Gradient Boosting Trees

CCP issues exist in many fields, including telecommunications (Calzada-Infante et al., 2020; Verbeke et al., 2014), banks and other financial institutions (Kaya et al., 2018; Keramati et al., 2016; Shirazi and Mohammadi, 2019; Van den Poel and Larivière, 2004), and higher education (e.g., student dropout) (Alkhasawneh and Hargraves, 2014; José et al., 2018). CCP is primarily tackled through binary classification using various methods such as decision tree (De Caigny et al., 2018), logistic regression

(Coussement et al., 2017), support vector machines (Chen et al., 2012), AdaBoost (Idris et al., 2019), and XGBT (Ahmad et al., 2019). The above models, which can be collectively referred to as machine learning methods, have been applied in multiple fields of management (Abedin et al., 2021; Ghosh et al., 2023; Sana et al., 2022; Yang et al., 2023), especially in the field of business management (Alfiero et al., 2022; Jiang et al., 2023; Li et al., 2014, 2020; Ouenniche et al., 2018).

XGBT is one of the most useful and commonly-used supervised learning technologies, which relies on gradient boosting principle to pursue better speed and performance (Ge et al., 2017). XGBT employs ensemble strategy to integrate several weak predictors and produces output. Relative to other gradient boosted trees methods, it involves a better regularization strategy that can effectively overcome overfitting and improve learning performance (Budak and Sarvari, 2021). Prior studies have explored the potential of XGBT for CCP. For example, Vo et al. (2021) used four supervised learning algorithms to conduct churn prediction based on structured and unstructured data. Pa & RB (Pa and RB, 2021) proposed a Fine-tuned XGBT method for customer churn prediction in telecom companies. The method solved the imbalance data, data sparsity and overfitting issues.

However, the importance of hyperparameter selection of XGBT has mostly been neglected in previous studies. Most hyperparameters of XGBT are either determined subjectively or by a brute force method such as *grid search* (GS) (Pandey and Shukla, 2021). However, for the above-mentioned hyperparameter selection methods, grid search requires subjectively defining candidate values of each hyperparameter considered. If there are too few candidate values of hyperparameters, the model might not reach its best possible performance. If there are too many candidate values, the computing resources consumed and required running times will grow exponentially. In addition, irrespective of the optimization scheme deployed to optimize XGBT's hyperparameters, statistical indicators are always employed to select the optimal hyperparameters. This process has recently received some criticisms because it ignored the primary objective of the learning process, that is, taking profitable action for clients through retention campaigns (Verbeke et al., 2017, 2012). To this end, goal-oriented indicators have been introduced by researchers to deal with the CCP problem and estimate the expected profit of retention campaigns (Hand, 2009; Verbeke et al., 2012). Profit indicators for CCP are designed to identify right target clients and consequently, formulate profitable retention campaigns.

In addition, in prior CCP studies research exploring XGBT, researchers only choose to optimize the hyperparameters that define its structure. Previous research has systematically ignored an important

parameter of XGBT which we hypothesize as important to maximize model profitability, i.e. the class weight coefficient. Thus, above-mentioned methods of hyperparameters selection cannot effectively deal with the number imbalance and misclassification cost imbalance of churners and non-churners in CCP research.

In this paper, we introduce BO-XGBT, which enhances XGBT by determining optimal parameters by BO that can construct the optimal XGBT structure with highest profit indicator value. Furthermore, the class weight coefficient of XGBT is also optimized by BO to alleviate the number imbalance and misclassification cost imbalance of churners and non-churners that are unique to CCP research. More relevant studies in terms of CCP based on XGBT are shown in Table 1.

Reference	Dataset Industry	Number of Datasets	Optimized Hyperparameters	Optimization Method	Optimization Objective Metric	Sensitivity Analysis	Running Time Analysis	Statistical Test
Duyen et al. (2017)	Telecom	One	No	No	No	No	No	No
Vo et al. (2018)	Financial	Four	No	No	No	No	No	No
Zhuang (2018)	E-commerce	One	No	No	No	No	No	No
Amin et al. (2019)	Telecom	Two	No	No	No	No	No	No
Ahmad et al. (2019)	Telecom	One	Number of decision trees (DTs)	GS	AUC	No	No	No
Tang et al. (2020)	Music and Telecom	Two	No	No	No	No	No	No
Al-Shatnwai and Faris (2020)	Telecom	One	Proportion of features adopted by DTs, gain threshold of node splitting, max depth of DTs, minimum sum of sample weights in children, and proportion of samples adopted by DTs	GS	Accuracy and F1 measure	No	No	No
Lalwani et al. (2021)	Telecom	One	Not disclosed	GS	AUC	No	No	No
Senthan et al. (2021)	Telecom	One	Learning rate, max_depth of DTs, minimum sum of sample weights in children, gain threshold of node splitting, and proportion of features adopted by DTs	GS	Accuracy	No	No	No
Vo et al. (2021)	Financial	Four	No	No	No	No	No	T-test
Pandey and Shukla (2021)	Bank	One	Proportion of features adopted by DTs, gain threshold of node splitting, max depth of DTs, and proportion of samples adopted by DTs	Sequential model-based optimization	Accuracy	No	No	No
Janssens et al. (2022)	Business-to-business	One	Boosting rounds, learning rate	GS	Expected maximum profit measure for B2B customer churn	Yes	No	No
The proposed BO-XGBT	Telecom, E-commerce, bank, newspaper, insurance, and uber	Eight	Learning rate, minimum sum of sample weights in children, max depth of DTs, class weight	BO	EMPC	Yes	Yes	Wilcoxon signed rank test

Table 1: Relevant studies on CCP based on XGBT. The column Statistical Test indicates which statistical test is deployed to compare methods in terms of predictive performance.

3 Methodology

In this section, we introduced the methodologies involved in our paper, including profit-driven evaluation of CCP models, XGBT, and profit-driven BO-XGBT. The methodology of our study, including the training process of BO-XGBT, is shown visually in Figure 1.

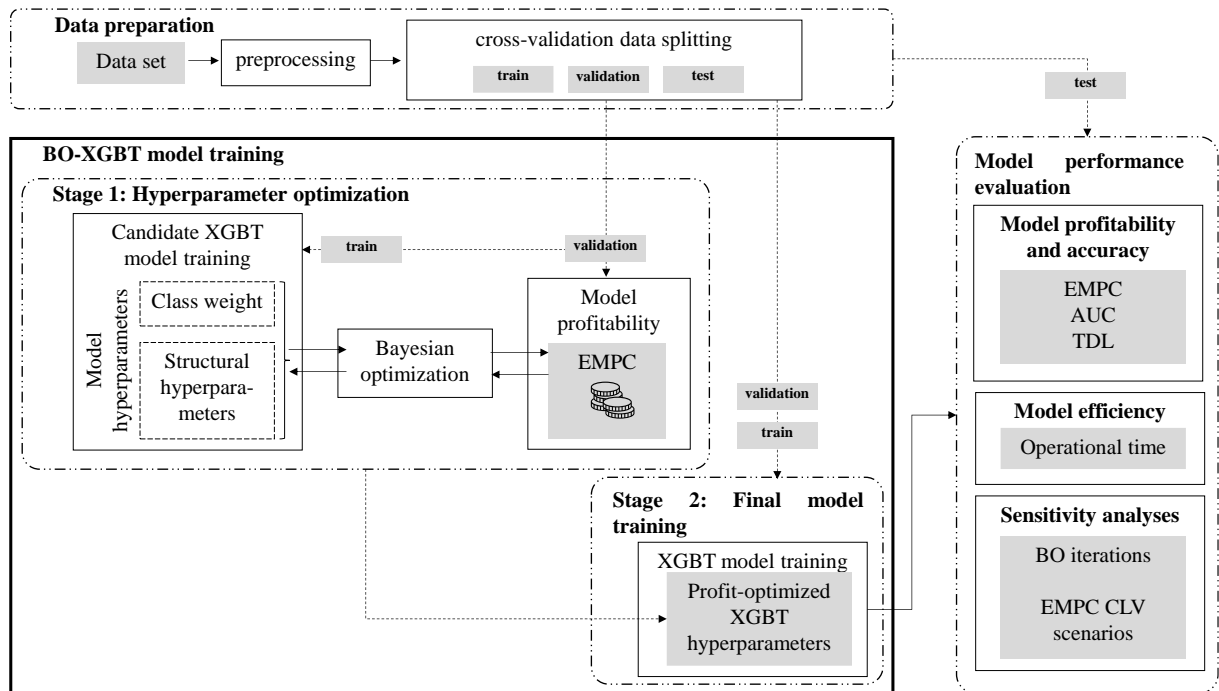


Figure 1: Graphical representation of the methodological framework for this study and of the model training process of BO-XGBT

3.1 Profit-Driven Evaluation of CCP Models

CCP can be regarded as a binary classification issue with the predicted binary label of $Y \in \{0,1\}$. Researchers used confusion matrix to present the classification results, whose definition is shown in Table 2. In this table, π_0 and π_1 are the prior probabilities of non-churners and churners, π_0 denotes $N/(M+N)$ and π_1 denotes $M/(M+N)$, respectively. F_0 and F_1 are the cumulative density functions of non-churner and churner, respectively. $F_0 = [FP/(FP+TN)] \times 100\%$ and $F_1 = [TP/(FN+TP)] \times 100\%$.

		Predicted	
		Churner	Non-Churner
Actual	Churner	TP ($\pi_1 \times F_1 \times (M + N)$)	FN ($\pi_1 \times (1 - F_1) \times (M + N)$)
	Non-Churner	FP ($\pi_0 \times F_0 \times (M + N)$)	TN ($\pi_0 \times (1 - F_0) \times (M + N)$)

Table 2: Definition of confusion matrix

Based on the confusion matrix, we can further calculate the performance metrics of classification models. One of the well accepted performance metrics is AUC. AUC is calculated without considering classification threshold, making it one of the most popular indicators to objectively test the classification ability. It depicts the effect of a predictor shown by ROC curve, which draws the true positive ratio and the false positive ratio under different threshold values (De Bock and Van den Poel, 2011). The mathematical expression of AUC is as follows:

$$AUC = \int TPR(t) [-FPR'(t)] dt \quad (1)$$

where TPR indicates true positive rate ($TPR = \frac{TP}{TP + FN}$), FPR indicates false positive rate ($FPR = \frac{FP}{FP + TN}$), and t indicates the classification threshold. The range of AUC is [0.5, 1], and larger AUC values represent better classification performance.

Decision-makers perform retention campaign on clients with highest churn risks based on classification outcomes. Correct classifications have benefit while misclassifications have cost. By comparing benefit and cost, we can further calculate profit. It must be noted that the misidentifying a churner as a non-churner (FN) may cause churner while misidentifying a non-churner as a churner (FP) also adds additional costs, such as the cost of contacting customers and incentive offer. To better measure the benefits of retaining customers and the costs of retention campaign, Verbraken et al. proposed the average classification profit function (Verbraken et al., 2013):

$$P_c(t; c, \overline{CLV}, \delta, \varphi) = \{ [c \cdot (1 - \delta) - \varphi] \cdot \pi_1 \cdot F_1 - (\delta + \varphi) \cdot \pi_0 \cdot F_0 \} \cdot \overline{CLV} \quad (2)$$

where t represents the classification threshold, \overline{CLV} represents the value of customer life value, c is the probability that churner is converted to non-churner. $\delta = d/\overline{CLV}$ and $\varphi = f/\overline{CLV}$, d is the cost of retaining a churner successfully, f is the cost of fighting for a churner. Based on $P_c(t; c, \overline{CLV}, \delta, \varphi)$, the EMPC indicator is then extended with the mathematical expression as (Verbraken et al., 2013):

$$EMPC = \int_c P_c(T(c); c, \overline{CLV}, \delta, \varphi) \cdot h(c) dc \quad (3)$$

where T represents the optimal classification threshold. $h(c)$ represents the probability density function

$$\text{of beta distribution for } c \text{ and } h(c)=h(c;\alpha,\beta)=\frac{c^{\alpha-1}(1-c)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1} du}.$$

Top-decile lift (TDL) is also an effective evaluation indicator for evaluating CCP models. It reveals the ratio of the percentage of actual churners in the top percent of the highest predicted churn probabilities, and the percentage of actual churners in all samples (De Bock and Van den Poel, 2011). TDL indicator stresses clients with the highest risk of attrition; thus, only a fraction of clients rather than the entirety is targeted by retention campaigns, and still, a majority of all potential churners can be effectively identified.

3.2 *Extreme Gradient Boosting Tree*

In the proposed BO-XGBT, XGBT, proposed by Chen, is adopted as the predictor (Chen and Guestrin, 2016). XGBT conducts a second-order Taylor expansion on the loss function using the gradient boosting decision tree method (Chehreh Chelgani et al., 2021). Relative to other gradient boosting techniques, XGBT could select a strong classification model among a series of poor classification models. The advantages of XGBT are as follows (Luo et al., 2021): (1) validly deal with missing values; (2) reduce overfitting; (3) parallel and distributed calculation enhance computational efficiency. Moreover, its objective is to use a gradient descent optimization algorithm and arbitrary differentiable loss functions to achieve the minimum of loss function via involving poor learners. XGBT attempts to minimize the objective function as follows (Trizoglou et al., 2021):

$$\begin{cases} \{\hat{\mathbf{y}}\} = \arg \min_{\{\hat{\mathbf{y}}\}} \text{Obj}_{XGBT}(\theta) = \sum_i L(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \\ f_k \in \Upsilon, \hat{\mathbf{y}} = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_i, \dots, \hat{y}_n\}, k = 1, 2, \dots, m \end{cases} \quad (4)$$

where L represents the training loss function, which is used to measure the deviation between the forecasted value \hat{y}_i and the original value y_i . f denotes a function in the functional space Υ , and Υ is the set of all probable regression trees. Ω denotes the regularization function used to depict the model complexity, and it can be used to reduce overfitting. The regularization $\Omega(f)$ is used to inflict punishment on the model complexity, which contributes to build simpler forecasting functions. The mathematical function of $\Omega(f)$ is as follows:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2 \quad (5)$$

where T is the number of leaves in each decision tree, γ is the complexity of a leaf, ω represent the vector scores of a leaf, and λ represents a parameter to measure the level of the penalty. More detailed operation mechanism of XGBT can be found in Chen & Guestrin (2016).

3.3 Profit-optimized XGBT

For implementing machine learning methods, model configuration and optimization are also significant (Liu and Lin, 2021). Therefore, we use BO to optimize the class weight and other hyperparameters of XGBT, so that XGBT with optimal hyperparameters based on training set can gain the most profit in validation set. To realize profit-driven objective, the objective function of BO is set as follows:

$$\begin{cases} \{\mathbf{Paras}\} = \arg \max_{\{\mathbf{Paras}\}} Obj_{BO} \\ Obj_{BO} = EMPC = \int_c P_C(T(c); c, \overline{CLV}, \delta, \varphi) \cdot h(c) dc \\ \overline{CLV} = \epsilon 200, d = \epsilon 10, f = \epsilon 1, \end{cases} \quad (6)$$

In Obj_{BO} , $P_C(T(c); c, \overline{CLV}, \delta, \varphi) = \{[c \cdot (1 - \delta) - \varphi] \cdot \pi_1 \cdot F_1 - (\delta + \varphi) \cdot \pi_0 \cdot F_0\} \cdot \overline{CLV}$, \overline{CLV} represents the average value of customer life value, and c denotes the probability that churner is converted to non-churner. d is the cost of remaining a churner successfully, f is the cost of fighting for a churner, and $\delta = d / \overline{CLV}$, $\varphi = f / \overline{CLV}$. π_0 denotes $N / (M + N)$ in the calculation of AUC , and π_1 denotes $M / (M + N)$ in the calculation of AUC . Besides, $F_0 = [FP / (FP + TN)] \times 100\%$, and $F_1 = [TP / (FN + TP)] \times 100\%$ (Verbraken et al., 2013), the definition of TP , FP , FN , and TN can be found in Table 2. In addition to $P_C(T(c); c, \overline{CLV}, \delta, \varphi)$, T represents the optimal classification threshold. $h(c)$ represents the probability density function of beta distribution, and

$$h(c) = h(c; \alpha, \beta) = \frac{c^{\alpha-1} (1-c)^{\beta-1}}{\int_0^1 u^{\alpha-1} (1-u)^{\beta-1} du}.$$

BO has the following advantages: (1) It considers previous parameter information and constantly updates the prior; (2) It has fewer iterations and higher speed of operation; (3) It is also robust against non-convex problems. Based on the above advantages of BO and the actual application scenarios of CCP, we choose BO as the optimizer of the proposed BO-XGBT. The search arranges of the structure

parameters of XGBT (\tilde{r}) are listed in Section 4.2. Besides, we set the search range of BO for class weight (\tilde{w}) as $[0, 50]$. The fitness function can be rewritten as:

$$\left\{ \begin{array}{l} \{\tilde{r}, \tilde{w}\} = \arg \max_{\{\tilde{r}, \tilde{w}\}} \int_c P_c(T(c); c, \overline{CLV}, \delta, \varphi) \cdot h(c) dc \\ s.t. \tilde{r} \text{ in certain range,} \\ \tilde{w} \in [0, 50], \overline{CLV} = \epsilon \cdot 200, d = \epsilon \cdot 10, f = \epsilon \cdot 1, \\ \delta = d / \overline{CLV}, \varphi = f / \overline{CLV}, \alpha = 6, \beta = 14, \\ h(c) = h(c; \alpha, \beta) = \frac{c^{\alpha-1} (1-c)^{\beta-1}}{\int_0^1 u^{\alpha-1} (1-u)^{\beta-1} du} \end{array} \right. \quad (7)$$

Figure A.1 in the Appendix contains detailed pseudocode of the full training process of BO-XGBT.

4 Experimental Validation

To verify the predictive performance and profitability of the proposed BO-XGBT, comprehensive experiments were conducted based on eight real-life churn prediction datasets. Section 4.1 presents the datasets and the detailed experimental settings are discussed in Section 4.2.

4.1 Datasets and Pre-Processing

Eight datasets are adopted to verify and benchmark the accuracy and profitability of the proposed BO-XGBT. This selection considers multiple contractual settings in sectors prominently featured in prior CCP research: telecom, e-commerce, bank, newspaper, insurance, and uber. The initial and the processed characteristics of these datasets before and after preprocessing are shown in Table 3.

The data sets are preprocessed in a number of ways in line with prior CCP literature. First, constant features are removed. Second, categorical features are one-hot encoded. Third, instances with missing values are removed.

Detailed Information	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Domain	E-Commerce ²	Telecom ³	Bank ⁴	Newspaper ⁵
Initial Sample Number	5630	7043	10000	15855
Initial Feature Number	19	32	13	18
Final Sample Number	5630	7040	10000	15850
Final Feature Number	35	26	12	61
Final Churn Rate	16.84%	26.55%	20.37%	19.15%
Detailed Information	Dataset 5	Dataset 6	Dataset 7	Dataset 8
Domain	Insurance ⁶	Uber ⁷	Telecom ⁸	Telecom ⁹
Initial Sample Number	33908	50000	51047	71047
Initial Feature Number	15	12	57	70
Final Sample Number	33900	50000	49770	69300
Final Feature Number	48	15	89	82
Final Churn Rate	11.70%	53.54%	28.64%	28.83%

Table 3: Dataset characteristics. We kindly refer to the reference URLs provided in footnotes for detailed feature description and additional information.

4.2 Experimental Setup

Twelve benchmark classifiers are considered as comparative models to illustrate the merit of the proposed BO-XGBT, including five individual classifiers, six ensemble classifiers, and the latest variant of XGBT for CCP. For individual classifiers, *Classification and Regression Tree* (CART) (Gordon et al., 1984), *K-Nearest Neighbor* (KNN) (Altman, 1992), *Least Absolute Shrinkage and Selection Operator* (LASSO) (Tishbirani, 1996), *Multi-Layer Perceptron* (MLP) (Haykin, 1999), and *Gaussian Naive Bayes Classifier* (GNBC) (Domingos and Pazzani, 1997). As a classical bagging-based ensemble model, we consider *Random Forest* (RF) (Breiman, 2001).

We compare BO-XGBT to a number of XGBT variants to confirm the importance of certain aspects of its design. First, BO-XGBT is compared to XGBT with hyperparameter optimization based on grid search (henceforward referred to as GS-XGBT). This will allow us to evaluate the added value of BO over a more conventional approach to hyperparameter tuning. Second, in order to highlight the importance of optimizing sample *class weight* (CW) in CCP, we compare BO-XGBT and GS-XGBT to their counterparts without class weight optimization. These variants are henceforward referred to as BO-XGBT (CW=1), and BO-XGBT (CW=balanced)), GS-XGBT (CW=1), and GS-XGBT

² <https://www.kaggle.com/datasets/ankitverma2010/ecommerce-customer-churn-analysis-and-prediction>

³ <https://www.kaggle.com/datasets/yeancz/telco-customer-churn-ibm-dataset>

⁴ <https://www.kaggle.com/mathchi/churn-for-bank-customers>

⁵ <https://www.kaggle.com/datasets/andieminoque/newspaper-churn>

⁶ <https://www.kaggle.com/datasets/mukul Singh/insurance-churn-prediction>

⁷ <https://www.kaggle.com/datasets/surbhivyas/uber-churn-rate-project>

⁸ <https://www.kaggle.com/datasets/jpacse/datasets-for-churn-telecom>

⁹ <https://www.kaggle.com/datasets/jpacse/telecom-churn-new-cell2cell-dataset>

(CW=balanced)) with. For BO-XGBT (CW=1) and GS-XGBT (CW=1), class weight is set as 1. For BO-XGBT (CW=balanced) and GS-XGBT (CW=balanced), considering that CCP problem is characterized by unbalanced misclassification cost and unbalanced sample numbers, we set the class weight deterministically as:

$$CW = \frac{N \times C_{churn}}{M \times C_{non-churn}} \quad (8)$$

where N and M denote the number of non-churner and churner, C_{churn} and $C_{non-churn}$ represent the cost of misidentifying a churner and non-churner.

$$C_{churn} = [c \cdot (1 - \delta) - \varphi] \cdot \pi_1 \cdot F_1 \cdot \overline{CLV} \quad (9)$$

$$C_{non-churn} = (\delta + \varphi) \cdot \pi_0 \cdot F_0 \cdot \overline{CLV} \quad (10)$$

where d is the cost of remaining a churner successfully, f is the cost of fighting for a churner, and $\delta = d / \overline{CLV}$, $\varphi = f / \overline{CLV}$. π_0 denotes $N / (M + N)$, and π_1 denotes $M / (M + N)$ in the calculation of *AUC*. Besides, $F_0 = [FP / (FP + TN)] \times 100\%$, and $F_1 = [TP / (FN + TP)] \times 100\%$ (Verbraken et al., 2013). In addition, we refer to the definition of average classification profit function in previous literature (Verbraken et al., 2013), and set c as 0.3.

Among several existing profit models, the latest proposed B2Boost (Janssens et al., 2022) can obtain more profit than other profit-driven models such as ProfLogit (Stripling et al., 2018), ProfTree (Höppner et al., 2020), and VerbrakenBoost¹⁰ (Verbraken et al., 2013). Therefore, we choose B2Boost as the benchmark model to compare with the proposed BO-XGBT, and adopt BO to select the optimal hyperparameters of B2Boost. For B2Boost, authors make a slight adjustment to the hyperparameters of EMPC for the application in B2B service industry. In this study, we focus on the general applicability of the proposed model across multiple industries, so in the B2Boost of our study, we use the default hyperparameters of EMPC in previous literature to make a fair comparison.

All these benchmark classifiers are implemented with hyperparameter optimization through Bayesian optimization with (100*number of optimized hyperparameters) iterations, similar to BO-

¹⁰ The mechanism of VerbrakenBoost is similar to GS-XGBT (CW=balanced) and BO-XGBT (CW=balanced).

XGBT. Candidate values and search range of hyperparameters are inspired by recent literature on CCP and shown in Table 4.

Model	Parameters	Parameter Type	Search Range or Candidate Values
BO-CART	max_features	Continuous numerical	[0.0001,1]
	min_samples_split	Discrete integer	[2, 3, ...,100]
	max_depth	Discrete integer	[3, 4, ...,10]
	min_samples_leaf	Discrete integer	[1, 2, ...,100]
BO-KNN	n_neighbors	Discrete integer	[1, 2, ...,100]
	leaf_size	Discrete integer	[1, 2, ...,100]
BO-LASSO	alpha	Continuous numerical	[0,1]
BO-MLP	max_iter	Discrete integer	50
	batch_size	Discrete integer	1024
	learning_rate_init	Continuous numerical	[0.0001,1]
	hidden_layer_sizes	Discrete integer	[50, 51, ...,250]
GNBC	-	-	-
BO-RF	n_estimators	Discrete integer	50
	min_samples_split	Discrete integer	[2, 3, ...,100]
	min_samples_leaf	Discrete integer	[1, 2, ...,100]
	max_depth	Discrete integer	[3, 4, ...,10]
GS-XGBT (CW=1), and GS-XGBT (CW=balanced)	n_estimators	Discrete integer	50
	learning_rate	Continuous numerical	[0.0001,0.25,0.5,0.75,1]
	min_child_weight	Discrete integer	[1,25,50,75,100]
	max_depth	Discrete integer	[3,5,7,9]
GS-XGBT	n_estimators	Discrete integer	50
	learning_rate	Continuous numerical	[0.0001,0.25,0.5,0.75,1]
	min_child_weight	Discrete integer	[1,25,50,75,100]
	max_depth	Discrete integer	[3,5,7,9]
	scale_pos_weight	Continuous numerical	[0.02,0.5,1,25,50]
BO-XGBT (CW=1), BO-XGBT (CW=balanced), and BO-B2Boost	n_estimators	Discrete integer	50
	learning_rate	Continuous numerical	[0.0001,1]
	min_child_weight	Discrete integer	[1, 2, ...,100]
	max_depth	Discrete integer	[3, 4, ...,10]
BO-XGBT	n_estimators	Discrete integer	50
	learning_rate	Continuous numerical	[0.0001,1]
	min_child_weight	Discrete integer	[1, 2, ...,100]
	max_depth	Discrete integer	[3, 4, ...,10]
	scale_pos_weight	Continuous numerical	[0.0001,50]
BO	iteration	Discrete integer	100*number of optimized hyperparameters
GS	cv	Discrete integer	5

Table 4: Candidate values and search range of hyperparameters. Parameter selection is based on the maximum EMPC value obtained in the validation set. Note: In this paper, only the more important hyperparameters in each model are searched or optimized. The remaining hyperparameters for each model are acquiescently set in the Scikit-Learn or XGBT packages.

The first aim of BO-XGBTs is to maximize CCP model profitability in terms of EMPC which is thus adopted as the primary criterion for hyperparameter tuning and for assessing model performance. In addition, we report AUC and TDL. To calculate EMPC estimates, relevant parameters are set based on previous studies (Stripling et al., 2018; Verbraken et al., 2013). Average value of customer life values (\overline{CLV}) is set as € 200, the probability that churner is converted to non-churner (c) is programmed to obey the Beta distribution, which α was set as 6 and β was set as 14. $\delta = d/\overline{CLV}$ and $\varphi = f/\overline{CLV}$, d is

the cost of remaining a churner successfully, f is the cost of fighting for a churner. d and f are set as € 10 and € 1, respectively.

To govern classifier training and report stable estimates of predictive performance, classifier training, parameter optimization and testing are replicated 10 times by means of a custom 10-fold cross-validation procedure. This procedure first involves the random division of the data sets into ten folds or parts of equal size. In each replication of the experiments, one fold serves as test set, while from the remaining 9 parts, one is randomly selected to serve as validation sample used for parameter tuning. Finally, to statistically compare classifiers in a pairwise fashion and over 8 data sets at once, we deploy the Wilcoxon signed rank test (Wilcoxon, 1945; Gattermann-Itschert & Thonemann, 2021). We apply this test on the 80 performance measures (8 data sets times 10 estimates obtained through cross-validation) obtained for each pair of classifiers.

5 Empirical Results

To verify the applicability of our proposed BO-XGBT in CCP, we conduct empirical analysis in three parts. First, the profitability of the proposed BO-XGBT is shown in Section 5.1, and then the high efficiency of Bayesian optimization for hyperparameter optimization is shown in Section 5.2. Subsequently, the robustness of BO-XGBT with respect to iteration numbers in Bayesian optimization, is listed in Section 5.3. Finally, Section 5.4 addresses the sensitivity of BO-XGBT with respect to

5.1 Predictive Performance and Profitability

Table 5-6 present the results of the proposed BO-XGBT and twelve comparative models in terms of profitability (EMPC) and predictive performance (TDL and AUC), while Table 7-8 present the results of statistical tests comparing BO-XGBT to benchmark methods in terms of EMPC, and comparing BO-XGBTs to GS-XGBTs in terms of EMPC specifically. All reported EMPC estimates are expressed in euros.

Dataset	Model	AUC	EMPC	TDL (10%)
# 1	BO-B2Boost	0.9907 (0.0046)	8.8878 (0.6894)	5.7920 (0.3779)
	BO-CART	0.9141 (0.0137)	6.9617 (0.7240)	4.8531 (0.3983)
	GNBC	0.7805 (0.0284)	4.6293 (0.6079)	3.0368 (0.3189)
	BO-KNN	0.8740 (0.0263)	6.9885 (0.7152)	4.7476 (0.3412)
	BO-LASSO	0.8812 (0.0236)	6.1063 (0.7336)	4.6182 (0.3431)
	BO-MLP	0.9394 (0.0176)	7.1660 (0.6781)	5.1032 (0.2654)
	BO-RF	0.9738 (0.0077)	8.0660 (0.7080)	5.7339 (0.3291)
	GS-XGBT (CW=1)	0.9924 (0.0047)	9.0027 (0.6854)	5.8245 (0.3936)
	BO-XGBT (CW=1)	0.9931 (0.0046)	9.0693 (0.6799)	5.8007 (0.3451)
	GS-XGBT (CW=balanced)	0.9944 (0.0038)	8.9187 (0.5783)	5.8985 (0.3612)
	BO-XGBT (CW=balanced)	0.9932 (0.0040)	9.0041 (0.6577)	5.8575 (0.3934)
	GS-XGBT	0.9940 (0.0032)	9.0143 (0.6187)	5.8362 (0.3917)
	BO-XGBT	0.9940 (0.0043)	9.1028 (0.6669)	5.9123 (0.4134)
	# 2	BO-B2Boost	0.8486 (0.0120)	10.5586 (0.6746)
BO-CART		0.8418 (0.0112)	10.4392 (0.6376)	2.8216 (0.1887)
GNBC		0.8424 (0.0116)	10.4348 (0.7567)	2.7475 (0.2181)
BO-KNN		0.8424 (0.0132)	10.3550 (0.7398)	2.7670 (0.2022)
BO-LASSO		0.8460 (0.0133)	10.4114 (0.7748)	2.7810 (0.2486)
BO-MLP		0.8443 (0.0222)	10.4814 (0.7085)	2.7942 (0.3223)
BO-RF		0.8565 (0.0122)	10.6393 (0.6221)	2.8700 (0.2113)
GS-XGBT (CW=1)		0.8585 (0.0126)	10.7293 (0.6991)	2.9199 (0.2377)
BO-XGBT (CW=1)		0.8565 (0.0123)	10.6676 (0.6876)	2.9253 (0.2386)
GS-XGBT (CW=balanced)		0.8580 (0.0125)	10.6295 (0.7483)	2.9410 (0.2288)
BO-XGBT (CW=balanced)		0.8534 (0.0120)	10.6339 (0.7287)	2.8475 (0.2184)
GS-XGBT		0.8601 (0.0137)	10.7472 (0.6898)	2.8917 (0.2132)
BO-XGBT		0.8544 (0.0114)	10.7162 (0.7233)	2.8606 (0.2676)
# 3		BO-B2Boost	0.8504 (0.0142)	7.0924 (0.5073)
	BO-CART	0.8362 (0.0122)	6.8735 (0.5721)	3.7736 (0.2019)
	GNBC	0.7848 (0.0113)	6.1127 (0.5292)	3.0872 (0.2075)
	BO-KNN	0.7669 (0.0189)	5.7223 (0.4333)	2.8273 (0.2070)
	BO-LASSO	0.7674 (0.0161)	5.8880 (0.5898)	2.7074 (0.2279)
	BO-MLP	0.8437 (0.0282)	6.9338 (0.5540)	3.8528 (0.3983)
	BO-RF	0.8598 (0.0118)	7.1586 (0.4465)	4.0245 (0.1432)
	GS-XGBT (CW=1)	0.8652 (0.0102)	7.2684 (0.4687)	4.0779 (0.1006)
	BO-XGBT (CW=1)	0.8638 (0.0099)	7.3184 (0.4897)	4.0623 (0.1014)
	GS-XGBT (CW=balanced)	0.8634 (0.0101)	7.2306 (0.4824)	4.0295 (0.1570)
	BO-XGBT (CW=balanced)	0.8611 (0.0111)	7.2535 (0.4751)	4.0091 (0.0946)
	GS-XGBT	0.8642 (0.0092)	7.2667 (0.4109)	4.0643 (0.1225)
	BO-XGBT	0.8650 (0.0103)	7.3825 (0.4763)	4.0405 (0.1750)
	# 4	BO-B2Boost	0.8511 (0.0203)	6.6229 (0.5727)
BO-CART		0.8321 (0.0137)	6.1211 (0.5279)	3.6452 (0.1523)
GNBC		0.5869 (0.0114)	3.6528 (0.5062)	1.2937 (0.2142)
BO-KNN		0.7663 (0.0166)	5.0760 (0.6146)	2.8318 (0.1884)
BO-LASSO		0.7684 (0.0113)	5.1130 (0.5645)	3.1994 (0.1185)
BO-MLP		0.8243 (0.0207)	5.9698 (0.6490)	3.3742 (0.1692)
BO-RF		0.8431 (0.0157)	6.2396 (0.6296)	3.5935 (0.1877)
GS-XGBT (CW=1)		0.8725 (0.0110)	6.8034 (0.5704)	3.8583 (0.2093)
BO-XGBT (CW=1)		0.8728 (0.0110)	6.8523 (0.5445)	3.7927 (0.1879)
GS-XGBT (CW=balanced)		0.8684 (0.0126)	6.7483 (0.5744)	3.7405 (0.1909)
BO-XGBT (CW=balanced)		0.8700 (0.0123)	6.7810 (0.5573)	3.7141 (0.2082)
GS-XGBT		0.8724 (0.0107)	6.8045 (0.5428)	3.7949 (0.1879)
BO-XGBT		0.8739 (0.0119)	6.8924 (0.5519)	3.7495 (0.1758)

Table 5: Average CCP metrics of the proposed BO-XGBT and twelve comparative models corresponding to 10 folds for Dataset 1-4. Note: The values in brackets are the standard deviations of each indicator value in 10-fold cross-validation. Best performance for each metric among all models is highlighted in bold. The unit of EMPC is €.

Dataset	Model	AUC	EMPC	TDL (10%)
# 5	BO-B2Boost	0.9217 (0.0063)	4.2820 (0.3306)	5.1922 (0.1532)
	BO-CART	0.8888 (0.0118)	3.7707 (0.3330)	4.9020 (0.1616)
	GNBC	0.8073 (0.0135)	2.6557 (0.2801)	3.9196 (0.1184)
	BO-KNN	0.7995 (0.0111)	2.5873 (0.2510)	3.9741 (0.2520)
	BO-LASSO	0.9069 (0.0062)	3.9509 (0.2572)	4.9485 (0.1327)
	BO-MLP	0.9180 (0.0045)	4.1894 (0.3426)	5.0846 (0.1385)
	BO-RF	0.9210 (0.0060)	4.2196 (0.2931)	5.0914 (0.1850)
	GS-XGBT (CW=1)	0.9326 (0.0057)	4.4319 (0.3139)	5.3342 (0.2496)
	BO-XGBT (CW=1)	0.9320 (0.0056)	4.4494 (0.3106)	5.3092 (0.1384)
	GS-XGBT (CW=balanced)	0.9260 (0.0045)	4.3032 (0.2777)	5.2115 (0.1436)
	BO-XGBT (CW=balanced)	0.9272 (0.0041)	4.3283 (0.3093)	5.2719 (0.1808)
	GS-XGBT	0.9326 (0.0056)	4.4357 (0.3056)	5.3478 (0.2057)
	BO-XGBT	0.9332 (0.0062)	4.4752 (0.3147)	5.3078 (0.2492)
# 6	BO-B2Boost	0.9135 (0.0103)	27.9752 (0.6240)	1.7569 (0.0660)
	BO-CART	0.9140 (0.0179)	27.8593 (0.6136)	1.8275 (0.0576)
	GNBC	0.7750 (0.0073)	25.2843 (0.6438)	1.4945 (0.0361)
	BO-KNN	0.8689 (0.0061)	26.8742 (0.6437)	1.6957 (0.0319)
	BO-LASSO	0.8043 (0.0065)	27.8643 (0.6110)	1.4208 (0.0302)
	BO-MLP	0.8817 (0.0238)	27.8616 (0.6123)	1.6872 (0.0907)
	BO-RF	0.9107 (0.0051)	27.8925 (0.6081)	1.7766 (0.0267)
	GS-XGBT (CW=1)	0.9463 (0.0023)	28.0133 (0.6087)	1.8659 (0.0343)
	BO-XGBT (CW=1)	0.9461 (0.0023)	28.0276 (0.6152)	1.8652 (0.0367)
	GS-XGBT (CW=balanced)	0.9431 (0.0029)	28.0119 (0.6068)	1.8655 (0.0346)
	BO-XGBT (CW=balanced)	0.9437 (0.0035)	28.0174 (0.6273)	1.8651 (0.0352)
	GS-XGBT	0.9464 (0.0028)	28.0148 (0.6058)	1.8655 (0.0344)
	BO-XGBT	0.9441 (0.0023)	28.0457 (0.6092)	1.8666 (0.0351)
# 7	BO-B2Boost	0.6524 (0.0086)	8.9232 (0.3863)	1.7137 (0.0666)
	BO-CART	0.6321 (0.0060)	8.7225 (0.4314)	1.5779 (0.0579)
	GNBC	0.5720 (0.0090)	8.3588 (0.4483)	1.3217 (0.0735)
	BO-KNN	0.5598 (0.0088)	8.3355 (0.4343)	1.3005 (0.0591)
	BO-LASSO	0.6159 (0.0102)	8.5822 (0.4175)	1.5197 (0.0586)
	BO-MLP	0.5962 (0.0339)	8.5233 (0.4017)	1.4252 (0.1161)
	BO-RF	0.6589 (0.0065)	8.9569 (0.4192)	1.7631 (0.0567)
	GS-XGBT (CW=1)	0.6719 (0.0054)	9.0055 (0.4053)	1.8673 (0.0688)
	BO-XGBT (CW=1)	0.6743 (0.0057)	9.0255 (0.4055)	1.8804 (0.0519)
	GS-XGBT (CW=balanced)	0.6689 (0.0049)	8.9675 (0.4186)	1.8436 (0.0634)
	BO-XGBT (CW=balanced)	0.6685 (0.0065)	8.9768 (0.4140)	1.8312 (0.0838)
	GS-XGBT	0.6764 (0.0049)	9.0648 (0.4127)	1.9373 (0.0512)
	BO-XGBT	0.6747 (0.0065)	9.0658 (0.4022)	1.8734 (0.0655)
# 8	BO-B2Boost	0.6543 (0.0062)	9.0320 (0.3890)	1.7084 (0.0667)
	BO-CART	0.6386 (0.0051)	8.8785 (0.3979)	1.6270 (0.0382)
	GNBC	0.5718 (0.0082)	8.4695 (0.4331)	1.3393 (0.0676)
	BO-KNN	0.5634 (0.0052)	8.4541 (0.4328)	1.3347 (0.0540)
	BO-LASSO	0.6166 (0.0068)	8.6946 (0.4221)	1.5169 (0.0526)
	BO-MLP	0.6033 (0.0351)	8.6603 (0.4219)	1.4667 (0.1546)
	BO-RF	0.6601 (0.0046)	9.0600 (0.3873)	1.7675 (0.0478)
	GS-XGBT (CW=1)	0.6749 (0.0048)	9.1327 (0.3688)	1.8494 (0.0483)
	BO-XGBT (CW=1)	0.6778 (0.0045)	9.1716 (0.3725)	1.8820 (0.0640)
	GS-XGBT (CW=balanced)	0.6698 (0.0028)	9.0958 (0.3739)	1.8189 (0.0633)
	BO-XGBT (CW=balanced)	0.6716 (0.0048)	9.1103 (0.3711)	1.8487 (0.0511)
	GS-XGBT	0.6777 (0.0046)	9.1589 (0.3800)	1.9058 (0.0502)
	BO-XGBT	0.6784 (0.0038)	9.1797 (0.3894)	1.8773 (0.0521)

Table 6: Average CCP metrics of the proposed BO-XGBT and twelve comparative models corresponding to 10 folds for Dataset 5-8. Note: The values in brackets are the standard deviations of each indicator value in 10-fold cross-validation. Best performance for each metric among all models is highlighted in bold. The unit of EMPC is €.

Role	Model	Average Ranks	p-value	Z-Statistic	
Control	BO-XGBT	1.6000	-	-	
	BO-B2Boost	6.9375	$3.9163*10^{-14}$	7.5637	
	BO-CART	9.5750	$9.8477*10^{-15}$	7.7412	
	GNBC	12.0375	$9.4830*10^{-15}$	7.7460	
	BO-KNN	11.9500	$7.8495*10^{-15}$	7.7700	
	BO-LASSO	10.7750	$1.3308*10^{-14}$	7.7028	
	Benchmarks	BO-MLP	9.4625	$2.3313*10^{-14}$	7.6309
		BO-RF	7.3500	$1.2157*10^{-13}$	7.4151
		GS-XGBT (CW=1)	4.0375	$2.7561*10^{-10}$	6.3119
		BO-XGBT (CW=1)	3.2250	$3.0244*10^{-10}$	6.2975
GS-XGBT (CW=balanced)		5.6375	$3.0914*10^{-13}$	7.2904	
BO-XGBT (CW=balanced)		4.8750	$5.8461*10^{-13}$	7.2040	
	GS-XGBT	3.3000	$5.5045*10^{-8}$	5.4342	

Table 7: Wilcoxon signed rank test results between the proposed BO-XGBT and all benchmark models in terms of EMPC. Note: Average ranks are calculated based on the EMPC of proposed BO-XGBT and all benchmark models. Model with lower average rank is identified as the model can obtain more profit. The degrees of freedom are 79 for all tests.

Model	Average Ranks	Role	p-value	Z-Statistic
BO-XGBT (CW=1)	2.9375	Control	-	-
GS-XGBT (CW=1)	3.7750	Benchmark	$1.4685*10^{-6}$	4.8155
BO-XGBT (CW=balanced)	4.4625	Control	-	-
GS-XGBT (CW=balanced)	5.0500	Benchmark	$5.4855*10^{-3}$	2.7771
BO-XGBT	1.4625	Control	-	-
GS-XGBT	3.0750	Benchmark	$5.5045*10^{-8}$	5.4342

Table 8: Wilcoxon signed rank test results between BO-XGBTs and GS-XGBTs in terms of EMPC. Note: Average ranks are calculated based on the EMPC of all XGBT models. Model with lower average rank is identified as the model can obtain more profit. The degrees of freedom are 79 for all tests.

From these results, the following conclusions emerge. First, XGBT models outperform all other methods in terms of profitability (EMPC) on all data sets. Since all methods, including benchmarks feature hyperparameter optimization through Bayesian optimization, this clearly indicates the added value of stochastic gradient boosting for profit-oriented customer churn prediction over other classifiers. Second, the superiority does not translate to other metrics, i.e., top-decile lift and AUC. We note that these metrics are not the focus of BO-XGBT nor our experimental validation and are merely provided for comparison purposes. Nevertheless, all XGBT variants (not only BO-XGBT and GS-XGBT, but also BO-XGBT (CW=1), GS-XGBT (CW=1), BO-XGBT (CW=balanced), and GS-XGBT (CW=balanced)) consistently outperform the other classifiers considered. Third, XGBT models based on Bayesian optimization consistently and significantly outperform XGBT models based on grid search in terms of EMPC, clearly showcasing the beneficial effect of Bayesian optimization in optimizing hyperparameters in XGBT models. Fourth, BO-XGBT with the class weight optimized by Bayesian optimization in addition to other hyperparameters outperforms BO-XGBT variants without the

optimization for class weights (including BO-XGBT (CW=1) and BO-XGBT (CW=balanced)). This indicates that effectively optimizing the class weight of XGBT can effectively adjust the unbalanced sample distribution and unbalanced misclassification cost in CCP. Finally, both the XGBT models without class weight (GS-XGBT (CW=1) and BO-XGBT (CW=1)) and the XGBT models with the class weight selected by grid search or Bayesian optimization (GS-XGBT and BO-XGBT) outperforms the XGBT models with balanced class weight in terms of EMPC, indicating that simply and deterministically setting class weight may degrade CCP performance due to inappropriate class weight values.

5.2 Prediction Efficiency Comparison

In this section, an analysis of computational costs is made with respect to the computational time required to train the proposed BO-XGBT and other XGBT models. The average running times are listed in Table 9. According to Table 9, BO-XGBT clearly demonstrate higher prediction efficiency in comparison to GS-XGBT, which is confirmed by the result of a Wilcoxon signed rank test given in Table 10. However, the prediction efficiency of BO-XGBT(CW=1) is lower than GS-XGBT(CW=1), because GS-XGBT(CW=1) has fewer candidate combinations of hyperparameters compared with GS-XGBT, thus achieving higher operating efficiency. The efficiency of Bayesian optimization will be greatly improved compared with grid search in the application scenarios with more hyperparameters. The efficiency gain of BO-XGBT is attributable to the drastic reduction of the number of hyperparametric evaluations. For example, the GS-XGBT for each training process is 500 ($5*5*4*5$ unique hyperparameter configurations), while BO-XGBT identifies an optimal configuration in only 400 iterations.

Dataset	GS-XGBT (CW=1)	GS-XGBT (CW=balanced)	GS-XGBT	BO-XGBT (CW=1)	BO-XGBT (CW=balanced)	BO-XGBT
# 1	39.4000 (0.4899)	40.5000 (0.5)	184.7000 (0.4583)	88.7000 (32.1031)	63.5000 (3.0414)	85.9000 (3.8846)
# 2	50.1000 (0.3000)	52.0000 (0.0000)	234.9000 (0.7000)	57.4000 (2.6533)	56.1000 (3.4191)	74.8000 (8.9420)
# 3	58.0000 (0.0000)	60.3000 (0.4583)	271.3000 (0.4583)	80.0000 (33.1904)	60.4000 (2.2891)	76.8000 (2.1354)
# 4	65.4000 (0.4899)	67.0000 (0.0000)	316.5000 (1.0247)	83.2000 (25.9569)	80.0000 (19.4062)	92.0000 (2.8284)
# 5	116.8000 (0.7483)	121.8000 (0.4000)	562.3000 (1.1000)	155.6000 (52.3358)	99.3000 (4.5398)	145.9000 (8.9828)
# 6	113.6000 (0.4899)	115.0000 (0.4472)	550.2000 (3.7094)	138.1000 (62.0507)	104.8000 (11.9147)	126.7000 (10.0305)
# 7	266.4000 (1.0198)	265.6000 (0.9165)	2464.8000 (1304.949)	218.7000 (50.1399)	185.7000 (18.3796)	257.5000 (63.5866)
# 8	401.8000 (0.7483)	402.1000 (0.8307)	2113.6000 (606.4258)	326.4000 (27.1411)	271.2000 (17.6624)	439.7000 (82.4367)

Table 9: Average operational time of the proposed BO-XGBT and three comparative models corresponding to 10 folds. Note: The values in brackets are the standard deviations of operational time in 10-fold cross-validation, and the unit is seconds.

Model	Average Ranks	Role	p-value	Z-Statistic
BO-XGBT (CW=1)	4.4500	Control	-	-
GS-XGBT (CW=1)	2.1000	Benchmark	0.1592	1.4077
BO-XGBT (CW=balanced)	2.0875	Control	-	-
GS-XGBT (CW=balanced)	2.9000	Benchmark	2.3928×10^{-03}	-3.0366
BO-XGBT	3.2125	Control	-	-
GS-XGBT	6.0000	Benchmark	7.8364×10^{-15}	-7.7702

Table 10: Wilcoxon signed rank test results between BO-XGBTs and GS-XGBTs in terms of operational time. Note: Average ranks are calculated based on the operational time of all XGBT models. Model with lower average rank is identified as the model have higher efficiency. The degree of freedom is 79.

5.3 Sensitivity of BO-XGBT with respect to BO iterations

In this section, we shed light on the robustness of Bayesian optimization in the proposed BO-XGBT. Specifically, we investigate the sensitivity of our results to of the number of iterations in Bayesian optimization, by varying this parameter from 10×4^{11} to 150×4 for the 8 data sets considered in the experiments so far. Figure 2 shows the average EMPC values as well as standard errors obtained through cross-validation. Based upon these results, we can confirm that the Bayesian optimization with different iteration numbers has different performance in the proposed BO-XGBT. The proposed BO-XGBT will obtain relatively better prediction performance when the iteration number of Bayesian optimization is relatively high. In addition, the prediction profit remains slightly volatile, indicating that the Bayesian optimization can converge effectively in this application scenario. Finally, but most importantly, when the iterations number is selected as $[10 \times 4, 20 \times 4, \dots, 150 \times 4]$, the EMPC value obtained by BO-XGBT

¹¹ 4 denotes the number of optimized hyperparameters.

does not fluctuate greatly, indicating that it is feasible to use efficient Bayesian optimization as the profit-driven hyperparameter selection method.

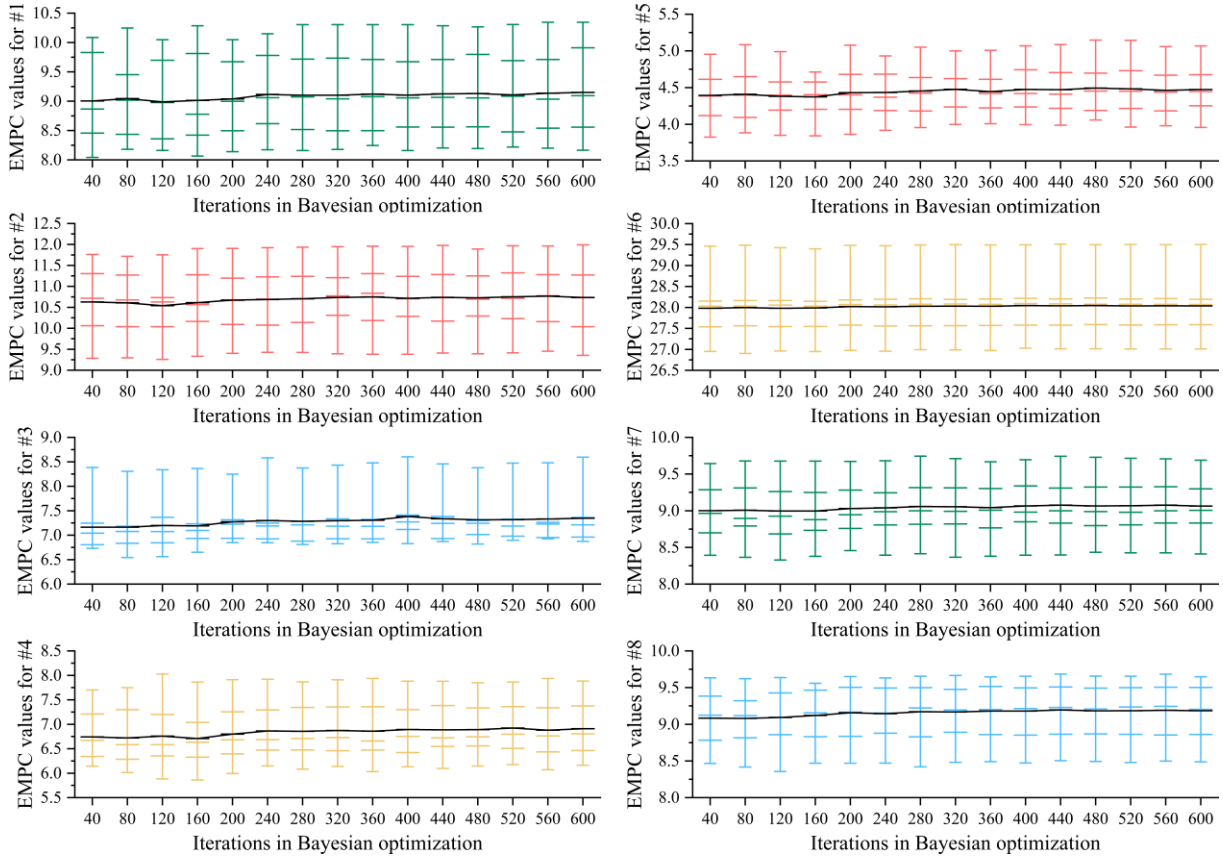


Figure 2: Average EMPC performance estimates and standard errors of the proposed BO-XGBT with varying iteration numbers

5.4 Sensitivity of of BO-XGBT with respect to EMPC return assumptions

To thoroughly assess the robustness of BO-XGBT within commercial scenarios, this section delves into a comprehensive exploration by configuring varying EMPC hyperparameters to emulate a spectrum of diverse business contexts. Additionally, an extensive comparative analysis is conducted between the proposed BO-XGBT model and a standard benchmark, shedding light on their respective profit-generating capacities. To achieve this, we introduce distinctive EMPC configurations, each with customer lifetime value (\overline{CLV}) settings of 100, 200, and 300, representing scenarios of low return ($\overline{CLV} = 100$), moderate return ($\overline{CLV} = 200$), and high return ($\overline{CLV} = 300$) of customer retention campaigns.

Dataset	Model	Low Return	Moderate Return	High Return
# 1	BO-B2Boost	3.6536 (0.4001)	8.8878 (0.6894)	13.9755 (1.0267)
	BO-CART	2.8165 (0.2613)	6.9617 (0.7240)	11.3006 (1.1956)
	GNBC	1.2905 (0.2721)	4.6293 (0.6079)	8.6404 (0.9750)
	BO-KNN	3.1098 (0.3492)	6.9885 (0.7152)	11.0293 (1.0196)
	BO-LASSO	2.2477 (0.4227)	6.1063 (0.7336)	10.4365 (1.0580)
	BO-MLP	2.9287 (0.3612)	7.1660 (0.6781)	11.7464 (1.0070)
	BO-RF	3.4135 (0.4426)	8.0660 (0.7080)	12.6501 (1.1091)
	GS-XGBT (CW=1)	4.0970 (0.3542)	9.0027 (0.6854)	13.9344 (1.0954)
	BO-XGBT (CW=1)	4.1260 (0.3659)	9.0693 (0.6799)	13.9875 (1.0803)
	GS-XGBT (CW=balanced)	4.0675 (0.3683)	8.9187 (0.5783)	13.9249 (1.1639)
	BO-XGBT (CW=balanced)	4.1036 (0.3277)	9.0041 (0.6577)	14.0851 (1.0867)
	GS-XGBT	4.0745 (0.3124)	9.0143 (0.6187)	13.9441 (1.0102)
	BO-XGBT	4.1751 (0.3529)	9.1028 (0.6669)	14.1757 (1.0912)
	# 2	BO-B2Boost	3.5729 (0.2623)	10.5586 (0.6746)
BO-CART		3.5639 (0.3494)	10.4392 (0.6376)	17.7906 (1.2392)
GNBC		3.6188 (0.3413)	10.4348 (0.7567)	17.8711 (1.2415)
BO-KNN		3.5584 (0.3304)	10.3550 (0.7398)	17.7614 (1.2763)
BO-LASSO		3.6469 (0.3524)	10.4114 (0.7748)	17.8343 (1.2086)
BO-MLP		2.6997 (1.4775)	10.4814 (0.7085)	17.3339 (1.8006)
BO-RF		3.7229 (0.3109)	10.6393 (0.6221)	18.0366 (1.0927)
GS-XGBT (CW=1)		3.7347 (0.2659)	10.7293 (0.6991)	18.1875 (1.2212)
BO-XGBT (CW=1)		3.7585 (0.3012)	10.6676 (0.6876)	18.2069 (1.1867)
GS-XGBT (CW=balanced)		3.7559 (0.3116)	10.6295 (0.7483)	18.1113 (1.2066)
BO-XGBT (CW=balanced)		3.7631 (0.2935)	10.6339 (0.7287)	18.2134 (1.1504)
GS-XGBT		3.7792 (0.3097)	10.7472 (0.6898)	18.1543 (1.1867)
BO-XGBT		3.7695 (0.2914)	10.7162 (0.7233)	18.2355 (1.2484)
# 3		BO-B2Boost	2.4772 (0.2100)	7.0924 (0.5073)
	BO-CART	2.3184 (0.1912)	6.8735 (0.5721)	11.8472 (0.6746)
	GNBC	1.7138 (0.2483)	6.1127 (0.5292)	11.3351 (0.8055)
	BO-KNN	1.5805 (0.2144)	5.7223 (0.4333)	10.7444 (0.7501)
	BO-LASSO	1.6133 (0.3234)	5.8880 (0.5898)	11.0045 (0.8170)
	BO-MLP	2.3670 (0.3917)	6.9338 (0.5540)	12.3403 (0.8156)
	BO-RF	2.6034 (0.1922)	7.1586 (0.4465)	12.3728 (0.8779)
	GS-XGBT (CW=1)	2.6352 (0.2071)	7.2684 (0.4687)	12.4940 (0.7442)
	BO-XGBT (CW=1)	2.6664 (0.1837)	7.3184 (0.4897)	12.5925 (0.8192)
	GS-XGBT (CW=balanced)	2.6631 (0.1981)	7.2306 (0.4824)	12.4168 (0.7327)
	BO-XGBT (CW=balanced)	2.6821 (0.2142)	7.2535 (0.4751)	12.4802 (0.7710)
	GS-XGBT	2.6264 (0.2175)	7.2667 (0.4109)	12.4829 (0.6435)
	BO-XGBT	2.7099 (0.1993)	7.3825 (0.4763)	12.6319 (0.8326)
	# 4	BO-B2Boost	2.1814 (0.2759)	6.6229 (0.5727)
BO-CART		2.0148 (0.2776)	6.1211 (0.5279)	10.9143 (0.8297)
GNBC		0.0976 (0.0468)	3.6528 (0.5062)	9.1005 (0.8473)
BO-KNN		1.3188 (0.2826)	5.0760 (0.6146)	9.9466 (0.9298)
BO-LASSO		1.5150 (0.2295)	5.1130 (0.5645)	9.7607 (0.8272)
BO-MLP		1.8970 (0.2420)	5.9698 (0.6490)	11.1029 (0.8533)
BO-RF		2.0626 (0.2967)	6.2396 (0.6296)	11.3220 (0.8956)
GS-XGBT (CW=1)		2.3708 (0.3166)	6.8034 (0.5704)	11.8542 (0.8332)
BO-XGBT (CW=1)		2.3866 (0.3110)	6.8523 (0.5445)	11.8810 (0.9274)
GS-XGBT (CW=balanced)		2.3384 (0.3049)	6.7483 (0.5744)	11.7401 (0.8802)
BO-XGBT (CW=balanced)		2.3645 (0.2925)	6.7810 (0.5573)	11.8629 (0.8231)
GS-XGBT		2.3660 (0.3083)	6.8045 (0.5428)	11.8441 (0.8315)
BO-XGBT		2.4301 (0.3176)	6.8924 (0.5519)	12.0321 (0.8392)

Table 11: Average EMPC metrics of the proposed BO-XGBT and twelve comparative models corresponding to 10 folds for Dataset 1-4. Note: The values in brackets are the standard deviations of each indicator value in 10-fold cross-validation. Best performance for each metric among all models is highlighted in bold. The unit of EMPC is €.

Dataset	Model	Low Return	Moderate Return	High Return
# 5	BO-B2Boost	1.3679 (0.1468)	4.2820 (0.3306)	7.4755 (0.4768)
	BO-CART	1.2354 (0.1459)	3.7707 (0.3330)	6.6565 (0.5135)
	GNBC	0.6884 (0.0923)	2.6557 (0.2801)	5.1435 (0.4913)
	BO-KNN	0.7288 (0.0846)	2.5873 (0.2510)	4.9404 (0.4722)
	BO-LASSO	1.2532 (0.1130)	3.9509 (0.2572)	6.9688 (0.4367)
	BO-MLP	1.3796 (0.1438)	4.1894 (0.3426)	7.2749 (0.5234)
	BO-RF	1.3762 (0.1416)	4.2196 (0.2931)	7.3265 (0.4572)
	GS-XGBT (CW=1)	1.5155 (0.1482)	4.4319 (0.3139)	7.6053 (0.5436)
	BO-XGBT (CW=1)	1.5466 (0.1479)	4.4494 (0.3106)	7.6197 (0.5483)
	GS-XGBT (CW=balanced)	1.4503 (0.1270)	4.3032 (0.2777)	7.4833 (0.4849)
	BO-XGBT (CW=balanced)	1.4864 (0.1473)	4.3283 (0.3093)	7.5379 (0.5186)
	GS-XGBT	1.5181 (0.1430)	4.4357 (0.3056)	7.5824 (0.5044)
	BO-XGBT	1.5539 (0.1427)	4.4752 (0.3147)	7.6577 (0.5377)
	# 6	BO-B2Boost	12.1392 (0.3222)	27.9752 (0.6240)
BO-CART		12.0273 (0.3379)	27.8593 (0.6136)	43.8532 (0.9826)
GNBC		9.9981 (0.3201)	25.2843 (0.6438)	41.0860 (1.0183)
BO-KNN		11.2534 (0.3317)	26.8742 (0.6437)	42.6303 (1.0156)
BO-LASSO		11.8970 (0.3219)	27.8643 (0.6110)	43.8457 (0.9645)
BO-MLP		11.9333 (0.3210)	27.8616 (0.6123)	43.8020 (0.9616)
BO-RF		12.0330 (0.3243)	27.8925 (0.6081)	43.8670 (0.9463)
GS-XGBT (CW=1)		12.1904 (0.3228)	28.0133 (0.6087)	43.9569 (0.9697)
BO-XGBT (CW=1)		12.1981 (0.3345)	28.0276 (0.6152)	43.9702 (0.9679)
GS-XGBT (CW=balanced)		12.1755 (0.3385)	28.0119 (0.6068)	43.9741 (0.9513)
BO-XGBT (CW=balanced)		12.1841 (0.3450)	28.0174 (0.6273)	43.9855 (0.9604)
GS-XGBT		12.1966 (0.3175)	28.0148 (0.6058)	43.9584 (0.9747)
BO-XGBT		12.2143 (0.3278)	28.0457 (0.6092)	43.9980 (0.9651)
# 7		BO-B2Boost	1.7453 (0.1602)	8.9232 (0.3863)
	BO-CART	1.6656 (0.1431)	8.7225 (0.4314)	16.9909 (0.6944)
	GNBC	1.1831 (0.1547)	8.3588 (0.4483)	16.8198 (0.7070)
	BO-KNN	1.1035 (0.1200)	8.3355 (0.4343)	16.8143 (0.7005)
	BO-LASSO	1.5170 (0.1688)	8.5822 (0.4175)	16.9016 (0.6899)
	BO-MLP	1.4606 (0.1863)	8.5233 (0.4017)	16.8621 (0.7343)
	BO-RF	1.8913 (0.1540)	8.9569 (0.4192)	17.1501 (0.6818)
	GS-XGBT (CW=1)	2.0210 (0.1765)	9.0055 (0.4053)	17.1727 (0.6678)
	BO-XGBT (CW=1)	2.0421 (0.1774)	9.0255 (0.4055)	17.1957 (0.6659)
	GS-XGBT (CW=balanced)	2.0098 (0.1728)	8.9675 (0.4186)	17.0814 (0.6846)
	BO-XGBT (CW=balanced)	2.0181 (0.1766)	8.9768 (0.4140)	17.1164 (0.6938)
	GS-XGBT	2.0405 (0.1728)	9.0648 (0.4127)	17.1952 (0.6577)
	BO-XGBT	2.0651 (0.1785)	9.0658 (0.4022)	17.2251 (0.6623)
	# 8	BO-B2Boost	1.8075 (0.1425)	9.0320 (0.3890)
BO-CART		1.7342 (0.1517)	8.8785 (0.3979)	17.1739 (0.6517)
GNBC		1.2105 (0.1614)	8.4695 (0.4331)	16.9974 (0.6821)
BO-KNN		1.1451 (0.1450)	8.4541 (0.4328)	16.9922 (0.6784)
BO-LASSO		1.5475 (0.1473)	8.6946 (0.4221)	17.0722 (0.6711)
BO-MLP		1.5541 (0.1744)	8.6603 (0.4219)	17.0550 (0.6810)
BO-RF		1.9346 (0.1443)	9.0600 (0.3873)	17.3084 (0.6296)
GS-XGBT (CW=1)		2.0796 (0.1225)	9.1327 (0.3688)	17.3345 (0.6209)
BO-XGBT (CW=1)		2.0809 (0.1244)	9.1716 (0.3725)	17.3670 (0.6357)
GS-XGBT (CW=balanced)		2.0539 (0.1218)	9.0958 (0.3739)	17.2469 (0.6471)
BO-XGBT (CW=balanced)		2.0626 (0.1065)	9.1103 (0.3711)	17.2785 (0.6365)
GS-XGBT		2.0935 (0.1193)	9.1589 (0.3800)	17.3545 (0.6355)
BO-XGBT		2.1110 (0.1300)	9.1797 (0.3894)	17.3909 (0.6488)

Table 12: Average CCP metrics of the proposed BO-XGBT and twelve comparative models corresponding to 10 folds for Dataset 5-8. Note: The values in brackets are the standard deviations of each indicator value in 10-fold cross-validation. Best performance for each metric among all models is highlighted in bold. The unit of EMPC is €.

Drawing insights from the EMPC values showcased in Tables 11 and 12, it becomes evident that across this spectrum of business contexts, the proposed BO-XGBT mostly yields notably elevated EMPC values. This robust performance reaffirms the pivotal role that BO-XGBT can play in augmenting profit potential for enterprises engaged in customer churn prediction. The outcomes from this analysis underscore the model's versatility and effectiveness, positioning it as a valuable asset for businesses seeking enhanced profitability through refined customer loss anticipation strategies.

6 Conclusion and Discussion

6.1 Main Conclusions

This study proposes BO-XGBT, a new profit-driven classifier for CCP based on extreme gradient boosting trees and Bayesian optimization of hyperparameters. Three experiments are performed based on eight CCP datasets obtained from multiple industries. From the experiments, we find that the profitability of the proposed BO-XGBT is significantly better than other benchmark models with higher EMPC values, which verified the superiority of the proposed BO-XGBT in providing profit maximization prediction results. Moreover, the prediction profitability of the proposed BO-XGBT based on Bayesian optimization is better than the GS-XGBT based on grid search in most scenes, indicating that using Bayesian optimization can make use of previous parameter information and improve CCP profitability. Moreover, the proposed BO-XGBT with the optimal class weight is superior to the BO-XGBTs with the default class weight and the subjectively setting class weight, indicating that the optimization of class weight has an additive contribution to CCP profitability. It is worth noting that the operational time of GS-XGBT is much longer than the proposed BO-XGBT. This reflects that, in the application scenarios with more hyperparameters, the Bayesian optimization which determines the optimal hyperparameter in an objective search form is more suitable compared with the grid search in a subjective search form for CCP. In addition, we performed sensitivity analysis on the iteration number of Bayesian optimization. The analysis results show that the profitability of the proposed BO-XGBT will fluctuate at a relatively lower level with relatively higher iteration number in Bayesian optimization. Finally, we adjusted the hyperparameters of the profit metric to simulate multiple business scenarios, and the experimental results show that the proposed model can gain significant advantages in multiple business contexts.

The main contributions can be concluded as follows: (1) The class weight and other hyperparameters of XGBT are optimized can perform more satisfactory results in CCP in terms of classification problems with imbalanced data and different misclassification costs. In this paper, the optimal class weight and other hyperparameters were combined to construct the profit-driven BO-XGBT model. The profit indicator was designed as the objective function. This strategy was inspired by previous studies that structured CCP models by profit indicators but not statistical metrics. The proposed BO-XGBT innovatively optimized the profit during the model training, which showed very positive prediction results. (2) The Bayesian optimization was combined with XGBT in CCP, which can realize efficient hyperparameter optimization on customer churn. Different from previous studies, the Bayesian optimization adopted in the proposed BO-XGBT determined the optimal class weight and other hyperparameters in an objective search form, which can ensure high efficiency for hyperparameter optimization. Therefore, the main application contribution is that the proposed profit-driven CCP model was applicative and effective in multiple industries.

6.2 Academic Implications

Our research has implications for both churn prediction model development and proactive customer retention management. We contribute to existing literature on churn prediction modeling in multiple fields. We develop a profit-optimized XGBT and offer new insights into which form is efficient in hyperparameters selection. The Bayesian optimization offers an efficient and adaptive approach for hyperparameter optimization. Its application to hyperparameter selection for XGBT can enhance CCP performance and yield effective solutions for analogous problems, thus advancing research and application of optimization algorithms within the field of business analysis. The integration of XGBT with the Bayesian optimization, employed for profit-driven customer churn prediction, underscores the significance of data-driven business decision-making. This integration holds paramount importance in achieving refined customer management and bolstering business competitiveness. The utilization of the proposed BO-XGBT for customer churn prediction not only yields accurate forecasting outcomes, but also furnishes enterprises with business intelligence and decision-making support. This confluence contributes to the progression of predictive analytics, generating positive impacts within practical business scenarios.

6.3 Managerial Implications

From a managerial perspective, our results also have important implications. By analyzing customer behaviors and trends, accurately predicting which customers are likely to churn enables enterprises to formulate targeted strategies, minimizing churn rates and enhancing sustainable business development.

Firstly, leveraging BO-XGBT-based CCP aids in optimizing resource allocation for businesses. By identifying high-value customers and those at risk of churning, companies can allocate resources strategically, investing more effort and capital in retaining the most valuable clientele. For instance, an online retailer can utilize churn prediction to identify customers with high purchase frequency but recent declines in activity, then entice them to continue buying through coupons or exclusive promotions.

Secondly, BO-XGBT-based CCP drives personalized customer communication and marketing strategies. Understanding individual customer needs and preferences empowers enterprises to send customized messages, boosting attraction and loyalty. For example, a fitness center can tailor workout plans based on users' exercise preferences and goals, sparking interest and mitigating churn risks.

Furthermore, BO-XGBT-based CCP contributes to product and service optimization. Analyzing feedback and reasons for churn among customers enables rapid issue identification and improvement. For instance, a restaurant chain identifying a high churn rate at a specific branch could conduct further investigations, potentially revealing declining service quality. Targeted employee training could then be implemented to enhance service quality.

Lastly, BO-XGBT-based CCP can play a pivotal role in maximizing long-term value. Timely intervention enables businesses to take proactive measures before customers exhibit signs of churn, fostering stronger customer relationships. For instance, if a passenger experienced multiple flight delays, an airline could proactively offer compensation and additional services to maintain loyalty and a long-term partnership. In conclusion, BO-XGBT-based CCP can serve as a robust management tool for businesses, offering key implications in reducing churn rates, resource allocation, personalized marketing, product enhancement, and long-term value maximization. These contributions facilitate continual business growth and development in highly competitive markets further destabilized by (recent) crises such as COVID-19, the Russia-Ukrainian war, climate change and environmental degradation

6.4 *Limitations and Future Research Directions*

In future research, the actual situation should be considered to measure the quantitative relationship between customer life value and the cost of retaining a cherner successfully, but it is more expensive to conduct actual investigations and then model for each case. In CCP modeling, the way to determine the quantitative relationship between customer life value and the cost of retaining a cherner successfully still needs further discussion. Besides, the proposed BO-XGBT can also be used in other fields where profit measures are involved, e.g., credit scoring and fraud detection. The primary adaptation of the proposed BO-XGBT to other problems would involve replacing the EMPC by appropriate domain-specific metrics.

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

- Abedin, M.Z., Moon, M.H., Hassan, M.K., Hajek, P., 2021. Deep learning-based exchange rate prediction during the COVID-19 pandemic. *Ann. Oper. Res.* <https://doi.org/10.1007/s10479-021-04420-6>
- Ahmad, A.K., Jafar, A., Aljoumaa, K., 2019. Customer churn prediction in telecom using machine learning in big data platform. *J. Big Data* 6, 1–24. <https://doi.org/10.1186/s40537-019-0191-6>
- Al-Shatnwai, A.M., Faris, M., 2020. Predicting customer retention using XGBoost and balancing methods. *Int. J. Adv. Comput. Sci. Appl.* 11, 704–712. <https://doi.org/10.14569/IJACSA.2020.0110785>
- Alfiero, S., Battisti, E., Hadjielias, E., 2022. Black box technology, usage-based insurance, and prediction of purchase behavior: Evidence from the auto insurance sector. *Technol. Forecast. Soc. Change* 183, 121896. <https://doi.org/10.1016/j.techfore.2022.121896>
- Alkhasawneh, R., Hargraves, R.H., 2014. --good--Developing a hybrid model to predict student first year retention and academic success in STEM disciplines using neural network. *J. STEM Educ. Innov. Res.* 15, 35–42.
- Altman, N.S., 1992. An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.* 46, 175–185. <https://doi.org/10.1080/00031305.1992.10475879>
- Amin, A., Shah, B., Khattak, A.M., Lopes Moreira, F.J., Ali, G., Rocha, A., Anwar, S., 2019. Cross-company customer churn prediction in telecommunication: A comparison of data transformation methods. *Int. J. Inf. Manage.* 46, 304–319. <https://doi.org/10.1016/j.ijinfomgt.2018.08.015>
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Budak, A., Sarvari, P.A., 2021. Profit margin prediction in sustainable road freight transportation using machine learning. *J. Clean. Prod.* 314, 127990. <https://doi.org/10.1016/j.jclepro.2021.127990>
- Calzada-Infante, L., Óskarsdóttir, M., Baesens, B., 2020. Evaluation of customer behavior with temporal centrality metrics for churn prediction of prepaid contracts. *Expert Syst. Appl.* 160, 113553. <https://doi.org/10.1016/j.eswa.2020.113553>

- Chehreh Chelgani, S., Nasiri, H., Alidokht, M., 2021. Interpretable modeling of metallurgical responses for an industrial coal column flotation circuit by XGBoost and SHAP-A “conscious-lab” development. *Int. J. Min. Sci. Technol.* 31, 1135–1144. <https://doi.org/10.1016/j.ijmst.2021.10.006>
- Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system, in: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. pp. 785–794. <https://doi.org/10.1145/2939672.2939785>
- Chen, Z.Y., Fan, Z.P., Sun, M., 2012. A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data. *Eur. J. Oper. Res.* 223, 461–472. <https://doi.org/10.1016/j.ejor.2012.06.040>
- Colgate, M., Stewart, K., Kinsella, R., 1996. Customer defection: a study of the student market in Ireland. *Int. J. Bank Mark.* 14, 23–29. <https://doi.org/10.1108/02652329610113144>
- Coussement, K., Lessmann, S., Verstraeten, G., 2017. A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry. *Decis. Support Syst.* 95, 27–36. <https://doi.org/10.1016/j.dss.2016.11.007>
- Dawes Farquhar, J., 2004. Customer retention in retail financial services: an employee perspective. *Int. J. Bank Mark.* 22, 86–99. <https://doi.org/10.1108/02652320410521700>
- De Bock, K.W., Van den Poel, D., 2011. An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction. *Expert Syst. Appl.* 38, 12293–12301. <https://doi.org/10.1016/j.eswa.2011.04.007>
- De Caigny, A., Coussement, K., De Bock, K.W., 2018. A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *Eur. J. Oper. Res.* 269, 760–772. <https://doi.org/10.1016/j.ejor.2018.02.009>
- De Caigny, A., Coussement, K., De Bock, K.W., Lessmann, S., 2020. Incorporating textual information in customer churn prediction models based on a convolutional neural network. *Int. J. Forecast.* 36, 1563–1578. <https://doi.org/10.1016/j.ijforecast.2019.03.029>
- Domingos, P., Pazzani, M., 1997. On the Optimality of the Simple Bayesian Classifier under Zero-One Loss. *Mach. Learn.* 29, 103–130. <https://doi.org/10.1023/a:1007413511361>
- Duyen, D., Phuc, H., Phuong, V., Tu, V., 2017. Customer Churn Prediction in an Internet Service Provider, in: *2017 IEEE International Conference on Big Data (Big Data)*. pp. 3928–3933. <https://doi.org/10.1109/BigData.2017.8258400>
- Gandomi, A., Zolfaghari, S., 2013. Profitability of loyalty reward programs: An analytical investigation. *Omega* 41, 797–807. <https://doi.org/10.1016/j.omega.2012.10.003>
- Ganesh, J., Arnold, M.J., Reynolds, K.E., 2000. Understanding the Customer Base of Service Providers: An Examination of the Differences between Switchers and Stayers. *J. Mark.* 64, 65–87. <https://doi.org/10.1509/jmkg.64.3.65.18028>
- Gattermann-Itschert, T., Thonemann, U.W., 2021. How training on multiple time slices improves performance in churn prediction. *Eur. J. Oper. Res.* 295, 664–674. <https://doi.org/10.1016/j.ejor.2021.05.035>
- Ge, Y., He, S., Xiong, J., Brown, D.E., 2017. Customer churn analysis for a software-as-a-service company, in: *2017 Systems and Information Engineering Design Symposium (SIEDS)*. <https://doi.org/10.1109/SIEDS.2017.7937698>
- Ghosh, I., Jana, R.K., Abedin, M.Z., 2023. An ensemble machine learning framework for Airbnb rental price modeling without using amenity-driven features. *Int. J. Contemp. Hosp. Manag.* <https://doi.org/10.1108/IJCHM-05-2022-0562>
- Gordon, A.D., Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. *Biometrics* 40, 874. <https://doi.org/10.2307/2530946>

- Hajek, P., Abedin, M.Z., Sivarajah, U., 2022. Fraud Detection in Mobile Payment Systems using an XGBoost-based Framework. *Inf. Syst. Front.* <https://doi.org/10.1007/s10796-022-10346-6>
- Hand, D.J., 2009. Measuring classifier performance: A coherent alternative to the area under the ROC curve. *Mach. Learn.* 77, 103–123. <https://doi.org/10.1007/s10994-009-5119-5>
- Haykin, S., 1999. *Neural Networks: A Comprehensive Foundation* (3rd Edition). *Knowl. Eng. Rev.* 13, 409–412. <https://doi.org/10.1017/S0269888998214044>
- Höppner, S., Stripling, E., Baesens, B., Broucke, S. vanden, Verdonck, T., 2020. Profit driven decision trees for churn prediction. *Eur. J. Oper. Res.* 284, 920–933. <https://doi.org/10.1016/j.ejor.2018.11.072>
- Idris, A., Iftikhar, A., Rehman, Z. ur, 2019. Intelligent churn prediction for telecom using GP-AdaBoost learning and PSO undersampling. *Cluster Comput.* 22, 7241–7255. <https://doi.org/10.1007/s10586-017-1154-3>
- Jabeur, S. Ben, Gharib, C., Mefteh-Wali, S., Arfi, W. Ben, 2021. CatBoost model and artificial intelligence techniques for corporate failure prediction. *Technol. Forecast. Soc. Change* 166, 120658. <https://doi.org/10.1016/j.techfore.2021.120658>
- Janssens, B., Bogaert, M., Bagué, A., Van den Poel, D., 2022. B2Boost: instance-dependent profit-driven modelling of B2B churn. *Ann. Oper. Res.* <https://doi.org/10.1007/s10479-022-04631-5>
- Jiang, C., Zhou, Y., Chen, B., 2023. Mining semantic features in patent text for financial distress prediction. *Technol. Forecast. Soc. Change* 190, 122450. <https://doi.org/10.1016/j.techfore.2023.122450>
- José, F., Bispo, M.D.S., Cássia, R. De, Pereira, D.F., 2018. Dropout and retention of undergraduate students in management : a study at a Brazilian Federal University. *RAUSP Manag. J.* 53, 74–85. <https://doi.org/10.1016/j.rauspm.2017.12.007>
- Kaya, E., Dong, X., Suhara, Y., Balcisoy, S., Bozkaya, B., Pentland, A. “Sandy,” 2018. Behavioral attributes and financial churn prediction. *EPJ Data Sci.* 7, 41. <https://doi.org/10.1140/epjds/s13688-018-0165-5>
- Keramati, A., Ghaneei, H., Mirmohammadi, S.M., 2016. Developing a prediction model for customer churn from electronic banking services using data mining. *Financ. Innov.* 2, 1–13. <https://doi.org/10.1186/s40854-016-0029-6>
- Lalwani, P., Mishra, M.K., Chadha, J.S., Sethi, P., 2021. Customer churn prediction system: a machine learning approach. *Computing* 104, 271–294. <https://doi.org/10.1007/s00607-021-00908-y>
- Lee, H., Lee, Y., Cho, H., Im, K., Kim, Y.S., 2011. Mining churning behaviors and developing retention strategies based on a partial least squares (PLS) model. *Decis. Support Syst.* 52, 207–216. <https://doi.org/10.1016/j.dss.2011.07.005>
- Li, G.-D., Masuda, S., Nagai, M., 2014. Predicting the subscribers of fixed-line and cellular phone in Japan by a novel prediction model. *Technol. Forecast. Soc. Change* 81, 321–330. <https://doi.org/10.1016/j.techfore.2013.05.004>
- Li, J.-P., Mirza, N., Rahat, B., Xiong, D., 2020. Machine learning and credit ratings prediction in the age of fourth industrial revolution. *Technol. Forecast. Soc. Change* 161, 120309. <https://doi.org/10.1016/j.techfore.2020.120309>
- Liu, X., Lin, Z., 2021. Impact of Covid-19 pandemic on electricity demand in the UK based on multivariate time series forecasting with Bidirectional Long Short Term Memory. *Energy* 227, 120455. <https://doi.org/10.1016/j.energy.2021.120455>
- Liu, Z., Jiang, P., Wang, J., Du, Z., Niu, X., Zhang, L., 2023. Hospitality order cancellation prediction from a profit-driven perspective. *Int. J. Contemp. Hosp. Manag.* 35, 2084–2112. <https://doi.org/10.1108/IJCHM-06-2022-0737>
- Lu, N., Lin, H., Lu, J., Zhang, G., 2014. A customer churn prediction model in telecom industry using boosting. *IEEE Trans. Ind. Informatics* 10, 1659–1665. <https://doi.org/10.1109/TII.2012.2224355>

- Luo, J., Zhang, Z., Fu, Y., Rao, F., 2021. Time series prediction of COVID-19 transmission in America using LSTM and XGBoost algorithms. *Results Phys.* 27, 104462. <https://doi.org/10.1016/j.rinp.2021.104462>
- Maldonado, S., Flores, Á., Verbraken, T., Baesens, B., Weber, R., 2015. Profit-based feature selection using support vector machines - General framework and an application for customer retention. *Appl. Soft Comput. J.* 35, 740–748. <https://doi.org/10.1016/j.asoc.2015.05.058>
- Maldonado, S., López, J., Vairetti, C., 2020. Profit-based churn prediction based on Minimax Probability Machines. *Eur. J. Oper. Res.* 284, 273–284. <https://doi.org/10.1016/j.ejor.2019.12.007>
- Nitzan, I., Libai, B., 2011. Social Effects on Customer Retention. *J. Mark.* 75, 24–38. <https://doi.org/10.1509/jm.10.0209>
- Óskarsdóttir, M., Bravo, C., Verbeke, W., Sarraute, C., Baesens, B., Vanthienen, J., 2017. Social network analytics for churn prediction in telco: Model building, evaluation and network architecture. *Expert Syst. Appl.* 85, 204–220. <https://doi.org/10.1016/j.eswa.2017.05.028>
- Ouenniche, J., Pérez-Gladish, B., Bouslah, K., 2018. An out-of-sample framework for TOPSIS-based classifiers with application in bankruptcy prediction. *Technol. Forecast. Soc. Change* 131, 111–116. <https://doi.org/10.1016/j.techfore.2017.05.034>
- Pa, S., RB, D., 2021. Customer churn prediction in telecommunication industry through machine learning based Fine-tuned XGBoost algorithm, in: *Proceedings of the International Conference on Innovative Computing & Communication (ICICC) 2021*. <https://doi.org/10.2139/ssrn.3835039>
- Pandey, A.S., Shukla, K.K., 2021. Application of Bayesian Automated Hyperparameter Tuning on Classifiers Predicting Customer Retention in Banking Industry, in: *Advances in Intelligent Systems and Computing*. pp. 83–100. https://doi.org/10.1007/978-981-15-5619-7_7
- Pekel Ozmen, E., Ozcan, T., 2022. A novel deep learning model based on convolutional neural networks for employee churn prediction. *J. Forecast.* 41, 539–550. <https://doi.org/10.1002/for.2827>
- Reinartz, W., Kumar, V., 2002. The mismanagement of customer loyalty. *Harv. Bus. Rev.* 80, 86-94,125.
- Sana, J.K., Abedin, M.Z., Rahman, M Sohel, Rahman, M Saifur, 2022. A novel customer churn prediction model for the telecommunication industry using data transformation methods and feature selection. *PLoS One* 17, e0278095. <https://doi.org/10.1371/journal.pone.0278095>
- Senthan, P., Rathnayaka, R.M.K.T., Kuhaneswaran, B., Kumara, B.T.G.S., 2021. Development of churn prediction model using XGBoost - Telecommunication Industry in Sri Lanka, in: *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*. pp. 1–7. <https://doi.org/10.1109/IEMTRONICS52119.2021.9422657>
- Shirazi, F., Mohammadi, M., 2019. A big data analytics model for customer churn prediction in the retiree segment. *Int. J. Inf. Manage.* 48, 238–253. <https://doi.org/10.1016/j.ijinfomgt.2018.10.005>
- Stripling, E., vanden Broucke, S., Antonio, K., Baesens, B., Snoeck, M., 2018. Profit maximizing logistic model for customer churn prediction using genetic algorithms. *Swarm Evol. Comput.* 40, 116–130. <https://doi.org/10.1016/j.swevo.2017.10.010>
- Tang, Q., Xia, G., Zhang, X., Long, F., 2020. A Customer Churn Prediction Model Based on XGBoost and MLP, in: *2020 International Conference on Computer Engineering and Application (ICCEA)*. pp. 608–612. <https://doi.org/10.1109/ICCEA50009.2020.00133>
- Tishbirani, R., 1996. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. Ser. B* 58, 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Trizoglou, P., Liu, X., Lin, Z., 2021. Fault detection by an ensemble framework of Extreme Gradient Boosting (XGBoost) in the operation of offshore wind turbines. *Renew. Energy* 179, 945–962. <https://doi.org/10.1016/j.renene.2021.07.085>

- Van den Poel, D., Larivière, B., 2004. Customer attrition analysis for financial services using proportional hazard models. *Eur. J. Oper. Res.* 157, 196–217. [https://doi.org/10.1016/S0377-2217\(03\)00069-9](https://doi.org/10.1016/S0377-2217(03)00069-9)
- Verbeke, W., Baesens, B., Bravo, C., 2017. Profit-Driven Model Evaluation and Implementation, Profit-Driven Business Analytics, Wiley Online Books. <https://doi.org/10.1002/9781119443179.ch6>
- Verbeke, W., Dejaeger, K., Martens, D., Hur, J., Baesens, B., 2012. New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *Eur. J. Oper. Res.* 218, 211–229. <https://doi.org/10.1016/j.ejor.2011.09.031>
- Verbeke, W., Martens, D., Baesens, B., 2014. Social network analysis for customer churn prediction. *Appl. Soft Comput.* 14, 431–446. <https://doi.org/10.1016/j.asoc.2013.09.017>
- Verbraken, T., Bravo, C., Weber, R., Baesens, B., 2014. Development and application of consumer credit scoring models using profit-based classification measures. *Eur. J. Oper. Res.* 238, 505–513. <https://doi.org/10.1016/j.ejor.2014.04.001>
- Verbraken, T., Verbeke, W., Baesens, B., 2013. A novel profit maximizing metric for measuring classification performance of customer churn prediction models. *IEEE Trans. Knowl. Data Eng.* 25, 961–973. <https://doi.org/10.1109/TKDE.2012.50>
- Vo, N.N.Y., Liu, S., Brownlow, J., Chu, C., Culbert, B., Xu, G., 2018. Client churn prediction with call log analysis, in: *International Conference on Database Systems for Advanced Applications*. pp. 752–763. https://doi.org/10.1007/978-3-319-91458-9_47
- Vo, N.N.Y., Liu, S., Li, X., Xu, G., 2021. Leveraging unstructured call log data for customer churn prediction. *Knowledge-Based Syst.* 212, 106586. <https://doi.org/10.1016/j.knosys.2020.106586>
- Wilcoxon, F., 1945. Individual Comparisons by Ranking Methods. *Biometrics Bull.* 1, 80. <https://doi.org/10.2307/3001968>
- Yang, C., Abedin, M.Z., Zhang, H., Weng, F., Hajek, P., 2023. An interpretable system for predicting the impact of COVID-19 government interventions on stock market sectors. *Ann. Oper. Res.* <https://doi.org/10.1007/s10479-023-05311-8>
- Zhuang, Y., 2018. Research on E-commerce Customer Churn Prediction Based on Improved Value Model and XG-Boost Algorithm. *Manag. Sci. Eng.* 12, 51–56. <https://doi.org/10.3968/10816>

Appendix

Abbreviations	Full Name
AUC	Area Under the ROC Curve
BO	Bayesian Optimization
BO-CART	CART with the hyperparameters determined by BO
BO-KNN	KNN with the hyperparameters determined by BO
BO-LASSO	LASSO with the hyperparameters determined by BO
BO-MLP	MLP with the hyperparameters determined by BO
BO-RF	RF with the hyperparameters determined by BO
BO-XGBT	XGBT with the hyperparameters determined by BO
CART	Classification and Regression Tree
CCP	Customer Churn Prediction
CRM	Customer Relationship Management
CW	Class Weight
DT	Decision Tree
EMPC	Expected Maximum Profit Criterion
GNBC	Gaussian Naive Bayes Classifier
GS	Grid Search
GS-XGBT	XGBT with the hyperparameters determined by GS
KNN	K-Nearest Neighbor
LASSO	Least Absolute Shrinkage and Selection Operator
MLP	Multi-Layer Perceptron
RF	Random Forest
TDL	Top-Decile Lift
XGBT	eXtreme Gradient Boosting Tree

Table A1: List of Abbreviations

Input

- D : a training data sample, $D = \{(x_i, y_i)\}_{i=1}^n$; $x_i \in X \subset R^p$; $y_i \in Y = \{0,1\}$
- V : a validation data sample with the same features as D
- $F(\cdot)$: the fitting function for an XGBT model
- Θ : the search space for a set of q XGBT hyperparameters
- $H(\cdot)$: the target XGBT classifier performance optimization function (EMPC)
- i_{max} : maximum number of iterations for Bayesian optimization (BO)
- $U(\cdot)$: the acquisition function for BO (*Probability of Improvement*)

Bayesian hyperparameter optimization phase

Initialize S ; $S = \emptyset$

For $i = 1, 2, \dots, i_{max}$, do:

- Select $\theta_i = \arg \max_{\theta \in \Theta} U(\theta; S)$
- Evaluate θ_i
 - Train $\hat{F}(\theta_i; D)$, an XGBT model configured using θ_i on D
 - $z_i = H(\hat{F}(\theta_i; V))$
- Update $S \leftarrow S \cup (\theta_i, z_i)$
- Check for exit criteria

Return optimal hyperparameter set $\theta^* = \arg \max_{\theta \in S} \{z | (\theta, z) \in S\}$

Final model training phase

Train $F^*(\theta^*; D \cup V)$, an XGBT model configured using θ^* on $D \cup V$

Return BO-XGBT classifier F^*

Figure A.1: BO-XGBT model training pseudocode